

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 25, 2004, 04:22:47 ; Search time 4479.5 Seconds
(without alignments)
16284.480 Million cell updates/sec

Title: US-09-912-559-1
Perfect score: 1683
Sequence: 1 agtttgccaggatgtctga.....aaagtgaagtgccttctaa 1683

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues
Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl.*

1: gb_ba.*
2: gb_htg.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vi.*
15: em_ba.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_mu.*
20: em_om.*
21: em_ov.*
22: em_pat.*
23: em_ph.*
24: em_pl.*
25: em_ro.*
26: em_sts.*
27: em_un.*
28: em_un.*
29: em_un.*
30: em_htg_hum.*
31: em_htg_inv.*
32: em_htg_other.*
33: em_htg_mus.*
34: em_htg_pln.*
35: em_htg_rod.*
36: em_htg_mam.*
37: em_htg_vrt.*
38: em_sy.*
39: em_htgo_hum.*
40: em_htgo_mus.*
41: em_htgo_other.*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1683	100.0	1683	6	AX383954	AX383954 Sequence
2	1683	100.0	1683	6	BD177602	BD177602 Detection
3	1683	100.0	2408	9	S83182	S83182 hyaluronan-
4	1683	100.0	3008	6	AX409021	AX409021 Sequence
5	1683	100.0	3008	9	HUMHGFAL	D49742 Human mRNA
6	1681.4	99.9	2251	9	BC031412	BC031412 Homo sapi
7	1679.8	99.8	1683	6	AX383955	AX383955 Sequence
8	1679.8	99.8	1683	6	BD177603	BD177603 Detection
9	1144.2	68.0	2190	10	AK128915	AK128915 Mus muscu
10	1050.2	62.4	2119	10	BC031775	BC031775 Mus muscu
11	261.2	15.5	188349	9	AL390197	AL390197 Human DNA
12	258	15.3	157750	2	AC006097	AC006097 Homo sapi
13	242.6	14.4	397	6	AX305372	AX305372 Sequence
14	167.4	9.9	204214	2	AC115771	AC115771 Mus muscu
15	161	9.6	224208	2	AC106236	AC106236 Rattus no
16	161	9.6	24249	2	AC131859	AC131859 Rattus no
17	133	7.9	2035	10	AF099017	AF099017 Mus muscu
18	133	7.9	2063	10	AF224724	AF224724 Mus muscu
19	131.4	7.8	2067	10	BC019376	BC019376 Mus muscu
20	117.6	7.0	2033	6	E07591	E07591 DNA encodin
21	117.6	7.0	2033	6	I15469	I15469 Sequence 14
22	117.6	7.0	2033	6	I69315	I69315 Sequence 14
23	117.6	7.0	2036	6	AX333070	AX333070 Sequence
24	117.6	7.0	2036	6	AX408965	AX408965 Sequence
25	117.6	7.0	2036	9	HUMHGFAP	D4012 Homo sapien
26	100.6	6.0	2375	4	SSUPAR	X85801 B.taurus mR
27	99.8	5.9	2228	4	BTUPA	X85801 B.taurus mR
28	99.8	5.9	2355	4	BOVUKPA	L03546 Bos taurus
29	96.2	5.7	970	6	E07615	E07615 DNA encodin
30	96.2	5.7	970	6	I15462	I15462 Sequence 3
31	96.2	5.7	970	6	I69308	I69308 Sequence 3
32	94.6	5.6	1454	10	RNUFAM	X65651 R.norvegicu
33	94.6	5.6	2366	10	RNUPLAC	X63434 R.norvegicu
34	93.2	5.5	2299	10	MMURKR	X02389 Mouse mRNa
35	90.6	5.4	2512	6	A19618	A19618 Mammalian c
36	90.6	5.4	2585	10	BC061565	BC061565 Rattus no
37	89	5.3	2445	10	RATPATISS	M23697 Rat tissue-
38	87.6	5.2	2245	4	VMPPPA1A	M63987 Desmodus ro
39	87.6	5.2	2252	6	A07242	A07242 Artificial
40	87.6	5.2	2252	6	A07250	A07250 Artificial
41	87.6	5.2	2266	4	VMPPPA1S	M63986 Desmodus ro
42	84.8	5.0	2130	4	VMPPPAB	M63989 Desmodus ro
43	84.8	5.0	2137	6	A07245	A07245 Artificial
44	81.2	4.8	1620	6	I08184	I08184 Sequence 4
45	81.2	4.8	1725	4	VMPTPA	J05082 Vampire bat

ALIGNMENTS

RESULT 1
AX383954
LOCUS AX383954 1683 bp DNA linear PAT 19-MAR-2002
DEFINITION Sequence 1 from Patent EP1182258.
ACCESSION AX383954
VERSION AX383954.1 GI:19577525
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1
AUTHORS Roemisch, J., Stoeck, H.A., Feusner, A., Lang, W., Weimer, T.,
Becker, M., Nerlich, C. and Muth-Naumann, G.
TITLE Mutants of the factor vii activating protease and methods for their

Pred. No. is the number of results predicted by chance to have a

detection
Patent: EP 1182258-A 1 27-FEB-2002;
Aventis Behring GmbH (DE)
Location/Qualifiers
1..1683
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

JOURNAL
FEATURES
source

ORIGIN

Query Match 100.0%; Score 1683; DB 6; Length 1683;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1683; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGTTTGCAGGATGCTGATCTCCATGTTCTGCTGTTAAATGGCTCTGGTGGGAAAGACA 60
Db 1 ATGTTTGCAGGATGCTGATCTCCATGTTCTGCTGTTAAATGGCTCTGGTGGGAAAGACA 60
Qy 61 GCCTGTGGGTTCTCCCTGATGTCCTTTATTTGGAAGCCCTGGACCCAGACTGGACCCCTGAC 120
Db 61 GCCTGTGGGTTCTCCCTGATGTCCTTTATTTGGAAGCCCTGGACCCAGACTGGACCCCTGAC 120
Qy 121 CAGTATGATTACAGCTACGAGATTATATCAGGAAGAGACACACAGTAGCACACTTACC 180
Db 121 CAGTATGATTACAGCTACGAGATTATATCAGGAAGAGACACACAGTAGCACACTTACC 180
Qy 181 CATGCTGAGATCCTGACTGTTACTACATGAGGACCAAGCTGATPCCATGCCAGCCCAAC 240
Db 181 CATGCTGAGATCCTGACTGTTACTACATGAGGACCAAGCTGATPCCATGCCAGCCCAAC 240
Qy 241 CCCTGTGACAGGTGGGACTGCTCTCCATGGAGACACTTACATGAGCTGCTG 300
Db 241 CCCTGTGACAGGTGGGACTGCTCTCCATGGAGACACTTACATGAGCTGCTG 300
Qy 301 GCTCTTTTCTCGGAATAAGTGTGAGAAAGTSCAAAATACGTGCAAGGACAAACCCATGT 360
Db 301 GCTCTTTTCTCGGAATAAGTGTGAGAAAGTSCAAAATACGTGCAAGGACAAACCCATGT 360
Qy 361 GGGGGGGCCATGCTCTATCCAGAGTCTCTCTACTACCTGCTGCTGTTGTAACAC 420
Db 361 GGGGGGGCCATGCTCTATCCAGAGTCTCTCTACTACCTGCTGCTGTTGTAACAC 420
Qy 421 CTTTACAGAGTCCAGCTGCTCCCAAGTGGTTCCTGTATGAGGACCAAAACCCCTGCCAG 480
Db 421 CTTTACAGAGTCCAGCTGCTCCCAAGTGGTTCCTGTATGAGGACCAAAACCCCTGCCAG 480
Qy 481 AATGGGGTACCTGCTCCCGGATTAAGCGGAGATCCAAAGTTCACTGTGCTGCTCCGAC 540
Db 481 AATGGGGTACCTGCTCCCGGATTAAGCGGAGATCCAAAGTTCACTGTGCTGCTCCGAC 540
Qy 541 CAGTTCAAGGGAAATTCGTGAATAGGTTCTGATGACTGCTATGTTGGCGATGGCTAC 600
Db 541 CAGTTCAAGGGAAATTCGTGAATAGGTTCTGATGACTGCTATGTTGGCGATGGCTAC 600
Qy 601 TCTTACCGAGGAAATGAATAGGACAGTCAACCCAGCATGCTGCTCTTACTGGAATCTCC 660
Db 601 TCTTACCGAGGAAATGAATAGGACAGTCAACCCAGCATGCTGCTCTTACTGGAATCTCC 660
Qy 661 CACTCTCTTCAGGAGATTAACATGTTTATGGAGATGCTGAAACCCATGGGATT 720
Db 661 CACTCTCTTCAGGAGATTAACATGTTTATGGAGATGCTGAAACCCATGGGATT 720
Qy 721 GGGGAACACAATTTCTGCGAAACCCAGATGCGGACGAAAGCCCTGGTGTGCTTTATTA 780
Db 721 GGGGAACACAATTTCTGCGAAACCCAGATGCGGACGAAAGCCCTGGTGTGCTTTATTA 780
Qy 781 GTTACCAATGACAGGTGAATGGGAATGCTGATGCTCAGCTGCTCAGCCGAGGAC 840
Db 781 GTTACCAATGACAGGTGAATGGGAATGCTGATGCTCAGCTGCTCAGCCGAGGAC 840
Qy 841 GTTGCTTACCCAGGAAAGCCCACTGAGCCATCAACCAAGTTCCGGGGTTTGACTCC 900
Db 841 GTTGCTTACCCAGGAAAGCCCACTGAGCCATCAACCAAGTTCCGGGGTTTGACTCC 900

Qy 901 TGTGGAAGACTGAGATACGAGAGGAGATCAAGAGAATCTATGGAGGCTTTAAGAGC 960
Db 901 TGTGGAAGACTGAGATACGAGAGGAGATCAAGAGAATCTATGGAGGCTTTAAGAGC 960
Qy 961 ACGCGGGCAAGCAACCCATGGCAGGCTCCCTCCAGTCTCTGCTGCTTGAACATCTCC 1020
Db 961 ACGCGGGCAAGCAACCCATGGCAGGCTCCCTCCAGTCTCTGCTGCTTGAACATCTCC 1020
Qy 1021 ATGCCCCAGGCGCACTTCTGTTGGGGGCTGATCCACCCCTGCTGGGTGCTCACTGCT 1080
Db 1021 ATGCCCCAGGCGCACTTCTGTTGGGGGCTGATCCACCCCTGCTGGGTGCTCACTGCT 1080
Qy 1081 GCCCACTGACCCAGACATAAAAAACCAACATCTAAAGTGGTCTAGGGGACGAGACCTG 1140
Db 1081 GCCCACTGACCCAGACATAAAAAACCAACATCTAAAGTGGTCTAGGGGACGAGACCTG 1140
Qy 1141 AAGAAAGAGAAATTTTCATGAGCAGAGCTTTAGGGTGGAGAGATATTCAAGTACAGCCAC 1200
Db 1141 AAGAAAGAGAAATTTTCATGAGCAGAGCTTTAGGGTGGAGAGATATTCAAGTACAGCCAC 1200
Qy 1201 TACAATGAAAGAGATGAGATTTCCCAATGATATTGCAATGCTCAAGTTAAAGCCAGTG 1260
Db 1201 TACAATGAAAGAGATGAGATTTCCCAATGATATTGCAATGCTCAAGTTAAAGCCAGTG 1260
Qy 1261 GATGGTCACTGCTCTAGAAATCCAAATACGTGAAGACTGCTGCTTGGCTGATGGTCC 1320
Db 1261 GATGGTCACTGCTCTAGAAATCCAAATACGTGAAGACTGCTGCTTGGCTGATGGTCC 1320
Qy 1321 TTTCCCTCTGGGAGTCAAGTCCCAATCTCTGGCTGGGGTGTACAGAAACAGGAAAGGG 1380
Db 1321 TTTCCCTCTGGGAGTCAAGTCCCAATCTCTGGCTGGGGTGTACAGAAACAGGAAAGGG 1380
Qy 1381 TCCCGGCGAGTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1440
Db 1381 TCCCGGCGAGTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1440
Qy 1441 CAACTCTATGACCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1500
Db 1441 CAACTCTATGACCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1500
Qy 1501 GGGCAAGACACCTGCGGAGTGAATCTGGAGGCCCCCTGACCTGTGAGAGGACGCGCAC 1560
Db 1501 GGGCAAGACACCTGCGGAGTGAATCTGGAGGCCCCCTGACCTGTGAGAGGACGCGCAC 1560
Qy 1561 TACTAGCTCTATGGGATGATGAGCTGGGGCTGGAGTGGGAAAGAGGCGGCTCTAC 1620
Db 1561 TACTAGCTCTATGGGATGATGAGCTGGGGCTGGAGTGGGAAAGAGGCGGCTCTAC 1620
Qy 1621 ACCCAAGTTACCAAAATTCCTGAATGGATCAAGCCACCATCAAAAGTGAAGTGGCTTC 1680
Db 1621 ACCCAAGTTACCAAAATTCCTGAATGGATCAAGCCACCATCAAAAGTGAAGTGGCTTC 1680
Qy 1681 TAA 1683
Db 1681 TAA 1683

RESULT 2

BD177602 1683 bp DNA linear PAT 16-APR-2003
LOCUS
DEFINITION Detection method with the use of factor VII activating protease
mutant and specific antibody.
ACCESSION BD177602
VERSION BD177602.1 GI:30014864
KEYWORDS JP 2002291486-A/1.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1683)
AUTHORS Roemisch, J., Stoeck, H.A., Feussner, A., Lang, W., Weimer, T.,
Becker, M., Nerlich, C. and Naumann, G.M.

TITLE Detection method with the use of factor VII activating protease
JOURNAL mutant and specific antibody
Patent: JP 2002291486-A 1 08-OCT-2002;
COMMENT AVENTIS BEHRING GMBH
OS Homo sapiens (human)
PN JP 2002291486-A/1
PD 08-OCT-2002
PF 25-JUL-2001 JP 2001224423
PR 26-JUL-2000 DE 10036641.4,10-OCT-2000 DE 10050040.4 PR
21-OCT-2000 DE 10052319.6,12-APR-2001 DE 10118706.8 PI
JUEGEN ROEMISCH, HANS ARNOLD STOEHR, ANNETTE FEUSSNER, WIEGAND PI
LANG,
PI THOMAS WEIMER, MARGRET BECKER, CLAUDIA NERLICH, GUDRUN MUTH PI
NAUMANN
PC C12N15/09, C07K16/40, C12N9/50, C12Q1/68, G01N33/53, G01N33/533, PC
G01N33/534,
PC G01N33/535//C12P21/08 C12N15/00
CC Detection method with the use of factor VII activating CC
protease mutant and
CC specific antibody
PH Key Location/Qualifiers
FT source 1. .1683
FT Location/Qualifiers
1. .1683
/organism='Homo sapiens (human)'.
/db_xref='taxon:9606'

ORIGIN

Query Match 100.0%; Score 1683; DB 6; Length 1683;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1683; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGTTTCCAGGATGCTGATCCCATGTTCTGCTGTTAATGGCTCTGGTGGGAAGACA 60
Db 1 ATGTTTCCAGGATGCTGATCCCATGTTCTGCTGTTAATGGCTCTGGTGGGAAGACA 60
QY 61 GCCTGGGTTCTCCCTGATGCTTTATTGGAAAGCTGGACCCAGACTGGACCCCTGAC 120
Db 61 GCCTGGGTTCTCCCTGATGCTTTATTGGAAAGCTGGACCCAGACTGGACCCCTGAC 120
QY 121 CAGTATGATTACAGCTACGAGGATTATATCAGGAAGAACACACAGTAGACACTTACC 180
Db 121 CAGTATGATTACAGCTACGAGGATTATATCAGGAAGAACACACAGTAGACACTTACC 180
QY 181 CATGCTGAGAACTCTGACTGGTACTACATGAGGACCAAGCTGATCCATGCCAGCCAC 240
Db 181 CATGCTGAGAACTCTGACTGGTACTACATGAGGACCAAGCTGATCCATGCCAGCCAC 240
QY 241 CCCTGTGAACACCGTGGGACTGCTCGTCCATGGGAGCACCTTACATGCGAGTGCCTG 300
Db 241 CCCTGTGAACACCGTGGGACTGCTCGTCCATGGGAGCACCTTACATGCGAGTGCCTG 300
QY 301 GCTCCTTTCTCGGGAATAAGTGTGAAAGTGCAGAAATACGTGGAAGACACACCATGT 360
Db 301 GCTCCTTTCTCGGGAATAAGTGTGAAAGTGCAGAAATACGTGGAAGACACACCATGT 360
QY 361 GGCGGGGCCAATGCTCTATTACCCAGAGTCCCTTACTACCGTGTGTCTGTAACAC 420
Db 361 GGCGGGGCCAATGCTCTATTACCCAGAGTCCCTTACTACCGTGTGTCTGTAACAC 420
QY 421 CCTTACACAGTCCCGACTGCTCCCAAGTGGTTCCTGTATGACGGCCAAACCCCTGCCAG 480
Db 421 CCTTACACAGTCCCGACTGCTCCCAAGTGGTTCCTGTATGACGGCCAAACCCCTGCCAG 480
QY 481 AATGGGCTACCTGCTCCCGCATAGCGAGATCCAGTTCCAGTGTGCTGCTGCCGAC 540
Db 481 AATGGGCTACCTGCTCCCGCATAGCGAGATCCAGTTCCAGTGTGCTGCTGCCGAC 540
QY 541 CAGTTCAAGGGGAAATCTGTGAATAGGTTCTGATGACTGCTATGTGGCGATGCTAC 600
Db 541 CAGTTCAAGGGGAAATCTGTGAATAGGTTCTGATGACTGCTATGTGGCGATGCTAC 600

QY 601 TCTTACCGAGGGAAAATGAATAGGACAGTCAACAGCATGCGTGCCTTTACTGGAACTCC 660
Db 601 TCTTACCGAGGGAAAATGAATAGGACAGTCAACAGCATGCGTGCCTTTACTGGAACTCC 660
QY 661 CACTCTCTTTCAGAGAAATTACACATGTTTATGGAGATGCTCAAAACCATGGAT 720
Db 661 CACTCTCTTTCAGAGAAATTACACATGTTTATGGAGATGCTCAAAACCATGGAT 720
QY 721 GGGGAACACAAATTTCTGCAGAAACCCAGATGCGGACGAAAGCCCTTGTCTTTATAA 780
Db 721 GGGGAACACAAATTTCTGCAGAAACCCAGATGCGGACGAAAGCCCTTGTCTTTATAA 780
QY 781 GTTACCAATGCAAGTGAATTCGGAAATCTGTGATGTTCTCAGCTGCTCAGCCAGGAC 840
Db 781 GTTACCAATGCAAGTGAATTCGGAAATCTGTGATGTTCTCAGCTGCTCAGCCAGGAC 840
QY 841 GTTGCCTTACCAGAGGAAACCCCACTGAGCCATCAACCAAGCTTCCGGGGTTTGACTCC 900
Db 841 GTTGCCTTACCAGAGGAAACCCCACTGAGCCATCAACCAAGCTTCCGGGGTTTGACTCC 900
QY 901 TGTGGAAGACTGAGATGAGAGAGAGATCAAGAGAAATCTATGGAGCTTTAAGAGC 960
Db 901 TGTGGAAGACTGAGATGAGAGAGAGAGATCAAGAGAAATCTATGGAGCTTTAAGAGC 960
QY 961 ACSCGGGCAAGCACCCATGCGAGGGCTCCCTCCAGTCCCTGCTGCTGCTGCTGCTGCT 1020
Db 961 ACSCGGGCAAGCACCCATGCGAGGGCTCCCTCCAGTCCCTGCTGCTGCTGCTGCTGCT 1020
QY 1021 ATGCCCCAGGCGCACTTCTGTGTTGGGGCGCTGATCCACCCCTGCTGCTGCTGCTGCT 1080
Db 1021 ATGCCCCAGGCGCACTTCTGTGTTGGGGCGCTGATCCACCCCTGCTGCTGCTGCTGCT 1080
QY 1081 GCCCAGCTGCAACGATGAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1140
Db 1081 GCCCAGCTGCAACGATGAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1140
QY 1141 AAGAAAGAGAAATTTCTGAGCAGAGCTTTAGGCTGGAGAGATATTCAGAGTACAGCCAC 1200
Db 1141 AAGAAAGAGAAATTTCTGAGCAGAGCTTTAGGCTGGAGAGATATTCAGAGTACAGCCAC 1200
QY 1201 TACAAAGAGAGATGAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1260
Db 1201 TACAAAGAGAGATGAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1260
QY 1261 GATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1320
Db 1261 GATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1320
QY 1321 TTTCCCTCTGGAGTGGAGTGCACATCTCTGGCTGGGGTGTACAGAAACAGGAAAGGG 1380
Db 1321 TTTCCCTCTGGAGTGGAGTGCACATCTCTGGCTGGGGTGTACAGAAACAGGAAAGGG 1380
QY 1381 TCCCGCAGCTCTCTGGATGCGAAAGTCAAGCTGATTCGCAACACTTTTGTGCAACTCCCGC 1440
Db 1381 TCCCGCAGCTCTCTGGATGCGAAAGTCAAGCTGATTCGCAACACTTTTGTGCAACTCCCGC 1440
QY 1441 CAATCTATGACCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1500
Db 1441 CAATCTATGACCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1500
QY 1501 GGGCAAGACACTGCGAGGGTACTCTGGAGGGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1560
Db 1501 GGGCAAGACACTGCGAGGGTACTCTGGAGGGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1560
QY 1561 TACTACGCTCTATGGGATAGTGGGGCTGGAGTGTGGGAGAGAGCCAGGGGTCTAC 1620
Db 1561 TACTACGCTCTATGGGATAGTGGGGCTGGAGTGTGGGAGAGAGCCAGGGGTCTAC 1620
QY 1621 ACCCAAGTACCAATTCCTGAATGGATCAAGCCCACTCAAAAGTGAAGTGGGCTTC 1680
Db 1621 ACCCAAGTACCAATTCCTGAATGGATCAAGCCCACTCAAAAGTGAAGTGGGCTTC 1680

Qy	1681 TAA 1683 1681 TAA 1683
Db	
RESULT 3	S83182 2408 bp mRNA linear PRI 11-FEB-1997
LOCUS	hyaluronan-binding protein-hepatocyte growth factor activator
DEFINITION	homolog [human, plasma, mRNA, 2408 nt].
ACCESSION	S83182
VERSION	S83182.1 GI:1836158
SOURCE	Homo sapiens (human)
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE	1. (bases 1 to 2408) Choi-Miura,N.H., Toke,T., Sumiya,J., Nakano,Y., Sano,Y., Mazda,T. and Tomita,M.
AUTHORS	Purification and characterization of a novel hyaluronan-binding protein (PHBP) from human plasma: it has three EGF, a kringle and a serine protease domain, similar to hepatocyte growth factor activator
TITLE	J. Biochem. 119 (6), 1157-1165 (1996)
JOURNAL	96425001
MEDLINE	8827452
PUBMED	entry [NCBI gibseq 179630] from the original journal article. This sequence comes from Fig. 3B.
REMARK	Location/Qualifiers
FEATURES	1..2408 <code>/organism="Homo sapiens"</code> <code>/mol_type="mRNA"</code> <code>/db_xref="taxon:9606"</code> <code>/gene="hyaluronan-binding protein, PHBP"</code> <code>/note="hepatocyte growth factor activator homolog; Method: conceptual translation with partial peptide sequencing; This sequence comes from Fig. 3B; PHBP"</code> <code>/codon_start=1</code> <code>/product="hyaluronan-binding protein"</code> <code>/protein_id="AAB46909.1"</code> <code>/db_xref="GI:1836159"</code>
gene	1..2408
CDS	97..1779 <code>/gene="hyaluronan-binding protein, PHBP"</code> <code>/note="hepatocyte growth factor activator homolog; Method: conceptual translation with partial peptide sequencing; This sequence comes from Fig. 3B; PHBP"</code> <code>/codon_start=1</code> <code>/product="hyaluronan-binding protein"</code> <code>/protein_id="AAB46909.1"</code> <code>/db_xref="GI:1836159"</code>
ORIGIN	Query Match 100.0%; Score 1683; DB 9; Length 2408; Best Local Similarity 100.0%; Pred. No. 0; Matches 1683; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy	1 ATGTTTCCAGGATGCTGATCTCATTTCATGAAGCCCTGGACCAGCTGCCCTTGAC 60
Db	97 ATGTTTCCAGGATGCTGATCTCATTTCATGAAGCCCTGGACCAGCTGCCCTTGAC 156
Qy	61 GCCTGTGGGTCTCCCTGATGCTTTATTGGAAGCCCTGGACCAGCTGCCCTTGAC 120
Db	157 GCCTGTGGGTCTCCCTGATGCTTTATTGGAAGCCCTGGACCAGCTGCCCTTGAC 216
Qy	121 CAGTATGATTAACGCTACGAGGATTTAATAACGGAAGAACAACAGTAGCACACTTACC 180
Db	217 CAGTATGATTAACGCTACGAGGATTTAATAACGGAAGAACAACAGTAGCACACTTACC 276

QY 1261 GATGGTCACTGTGCTCTAGAAATCCAAATACGTGAGACGTGTGTGCTTGCCTGATGGGTCC 1320
DB 1357 GATGGTCACTGTGCTCTAGAAATCCAAATACGTGAGACGTGTGTGCTTGCCTGATGGGTCC 1416
QY 1321 TTTCCCTCTGGGAGTGAGTGGCCACATCTCTGGCTGGGGTGTACAGAAAACAGAAAAGGG 1380
DB 1417 TTTCCCTCTGGGAGTGAGTGGCCACATCTCTGGCTGGGGTGTACAGAAAACAGAAAAGGG 1476
QY 1381 TCCCGGACCTCTGTGATGCCAAAGTCAAGCTGATTCGCAACACATTTGTGCAACTCCGCG 1440
DB 1477 TCCCGGACCTCTGTGATGCCAAAGTCAAGCTGATTCGCAACACATTTGTGCAACTCCGCG 1536
QY 1441 CAACTCTATGACACATGATGATGACAGTATGATCTGTGCAAGGAAATCTTCAGAAAACCT 1500
DB 1537 CAACTCTATGACACATGATGATGACAGTATGATCTGTGCAAGGAAATCTTCAGAAAACCT 1596
QY 1501 GGGCAAGACACCTGCCAGGGTGACTCTGGAGGCCCTGACTGTGAGAGACGGCACC 1560
DB 1597 GGGCAAGACACCTGCCAGGGTGACTCTGGAGGCCCTGACTGTGAGAGACGGCACC 1656
QY 1561 TACTAGCTCTATGGGATAGTGGCTGGGCTGGAGTGGGAGAGAGGCCAGGGGTCTAC 1620
DB 1657 TACTAGCTCTATGGGATAGTGGCTGGGCTGGAGTGGGAGAGAGGCCAGGGGTCTAC 1716
QY 1621 ACCAAGTACCAAAATCTGAAATGGATCAAGCCACCATCAAAAGTGAAGTGGCTTC 1680
DB 1717 ACCAAGTACCAAAATCTGAAATGGATCAAGCCACCATCAAAAGTGAAGTGGCTTC 1776
QY 1681 TAA 1683
DB 1777 TAA 1779

RESULT 4
LOCUS AX409021 3008 bp DNA linear PAT 14-JUN-2002
DEFINITION Sequence 1668 from Patent WO0229103.
ACCESSION AX409021
VERSION AX409021.1 GI:21441726
KEYWORDS Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Alvares, C., Horne, D., Peres-da-Silva, S. and Vockley, J. G.
TITLE Gene expression profiles in liver cancer
JOURNAL Patent: WO 0229103-A 1668 11-APR-2002;
GENE LOGIC INC (US)
FEATURES
source 1..3008
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
/note="EMBL/GenBank Accession No. D49742"

ORIGIN
Query Match 100.0%; Score 1683; DB 6; Length 3008;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1683; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGTTTGGCCAGGATGCTGATCTCCATGTTCTGCTGTTAATGGCTCTGCTGGGAAAGACA 60
DB 97 ATGTTTGGCCAGGATGCTGATCTCCATGTTCTGCTGTTAATGGCTCTGCTGGGAAAGACA 156
QY 61 GCCTGTGGGTCTCCCTGATGCTTTTATGGAAAGCTGGACCCAGACTGGACCCCTGAC 120
DB 157 GCCTGTGGGTCTCCCTGATGCTTTTATGGAAAGCTGGACCCAGACTGGACCCCTGAC 216
QY 121 CAGTATGATTACAGCTACGAGGATTATATCAGGAGAAACACCAAGTAGCACACTTACC 180
DB 217 CAGTATGATTACAGCTACGAGGATTATATCAGGAGAAACACCAAGTAGCACACTTACC 276

QY 181 CATGCTCAGAAATCCTGACTGGTACTACTACGTAGAGCCAGCTGATCCATGCCAGCCCAAC 240
DB 277 CATGCTCAGAAATCCTGACTGGTACTACTACGTAGAGCCAGCTGATCCATGCCAGCCCAAC 336
QY 241 CCCTGTGAACACAGGTGGGAGTCCCTCGTCCATGGGAGCACCTTCACATCAGCTCCCTG 300
DB 337 CCCTGTGAACACAGGTGGGAGTCCCTCGTCCATGGGAGCACCTTCACATCAGCTCCCTG 396
QY 301 GCTCCTTTCTCTGGGATAGTGTCAAGAGTCCAAATACGTGCAAGGACCAACCCATGT 360
DB 397 GCTCCTTTCTCTGGGAAATAGTGTCAAGAGTCCAAATACGTGCAAGGACCAACCCATGT 456
QY 361 GGGCGGGCCAAATGTCTCATTAACCCAGAGTCTCCCTACTACCCGTGTGTCTTAAACAC 420
DB 457 GGGCGGGCCAAATGTCTCATTAACCCAGAGTCTCCCTACTACCCGTGTGTCTTAAACAC 516
QY 421 CTTTACACAGGTCCAGCTCCTCCCAAGTGGTTCCTGTATGAGGCGCAACCCCTGCCAG 480
DB 517 CTTTACACAGGTCCAGCTCCTCCCAAGTGGTTCCTGTATGAGGCGCAACCCCTGCCAG 576
QY 481 AATGGGCTACCTGCTCCCGGCATAGCGGAGATCCAAAGTTCACCTGTGCTGTCCCGAC 540
DB 577 AATGGGCTACCTGCTCCCGGCATAGCGGAGATCCAAAGTTCACCTGTGCTGTCCCGAC 636
QY 541 CAGTTCAAGGGGAAATTTCTGTGAATAGTCTGTGATGACTGTCTATGTTGGCGATGGCTAC 600
DB 637 CAGTTCAAGGGGAAATTTCTGTGAATAGTCTGTGATGACTGTCTATGTTGGCGATGGCTAC 696
QY 601 TCTTACCGAGGAAATGAATAGACAGTCAACAGCATGCGTGCCTTTACTGCGAACTCC 660
DB 697 TCTTACCGAGGAAATGAATAGACAGTCAACAGCATGCGTGCCTTTACTGCGAACTCC 756
QY 661 CACCTCCTCTGCAAGGAAATTAACAATTTTATGAGGATGCTGAAACCCATGGGATT 720
DB 757 CACCTCCTCTGCAAGGAAATTAACAATTTTATGAGGATGCTGAAACCCATGGGATT 816
QY 721 GGGGAAACACAAATTTCTGCAGAACCCAGATGCGGAGAAAGCCCTGCTTTATTAA 780
DB 817 GGGGAAACACAAATTTCTGCAGAACCCAGATGCGGAGAAAGCCCTGCTTTATTAA 876
QY 781 GTTACCAATGACAGGTGAAATGGGAAATCTGTGATGCTCAGCTGCTCAGCCCAAGGAC 840
DB 877 GTTACCAATGACAGGTGAAATGGGAAATCTGTGATGCTCAGCTGCTCAGCCCAAGGAC 936
QY 841 GTTCCCTACCCAGGAAAGCCCACTGAGCCATCAACCAAGCTTCGGGGTTTGACTCC 900
DB 937 GTTCCCTACCCAGGAAAGCCCACTGAGCCATCAACCAAGCTTCGGGGTTTGACTCC 996
QY 901 TGTGAAAGACTGAGATAGCAGAGAGGAAAGATCAAGAGAAATCTATGGAGCTTTAAGAGC 960
DB 997 TGTGAAAGACTGAGATAGCAGAGAGGAAAGATCAAGAGAAATCTATGGAGCTTTAAGAGC 1056
QY 961 ACGCGGGCAAGCAACCATGGCAGGGTCCCTCAGTCTCGCTGCTGCTGACCATCTCC 1020
DB 1057 ACGCGGGCAAGCAACCATGGCAGGGTCCCTCAGTCTCGCTGCTGCTGACCATCTCC 1116
QY 1021 ATGCCCCAGGGCCACTTTCTGTGGGGCGCTGATCCACCCCTGCTGGGTGCTCACTGCT 1080
DB 1117 ATGCCCCAGGGCCACTTTCTGTGGGGCGCTGATCCACCCCTGCTGGGTGCTCACTGCT 1176
QY 1081 GCCCACTGCCAGCATATAAACCAGACATCTTAAAGTGGTGTGCTAGGGGACCGAGGACTG 1140
DB 1177 GCCCACTGCCAGCATATAAACCAGACATCTTAAAGTGGTGTGCTAGGGGACCGAGGACTG 1236
QY 1141 AAGAAAGAGAAATTTCTATGAGCAGAGCTTTAGGGTGGAGAGATATTCAAGTAGTACAGCCAC 1200
DB 1237 AAGAAAGAGAAATTTCTATGAGCAGAGCTTTAGGGTGGAGAGATATTCAAGTAGTACAGCCAC 1296
QY 1201 TACAATGAAGAGATGAGATTCCCAACATGATATTGATGCTCAAGTTGCTCAAGTTAAGCCAGTG 1260
DB 1297 TACAATGAAGAGATGAGATTCCCAACATGATATTGATGCTCAAGTTAAGTTAAGCCAGTG 1356
QY 1261 GATGGTCACTGTGCTCTAGAAATCCAAATACGTGAGACTGTGTGCTTGCCTGATGGGTCC 1320

[illegible]

457	GC	GGGGCCAA	TGCTCTCAT	TACCCAGAGT	CCTCCCTACT	ACCGCTGTCTGTGTAACAC	516
421	CT	TACACAGGT	CCCACTGCT	CCCCAAGTGGTTC	TGTTATG	CAGGCCAAACCCCTGCAG	480
517	CT	TACACAGGT	CCCACTGCT	CCCCAAGTGGTTC	TGTTATG	CAGGCCAAACCCCTGCAG	576
481	AT	GGGGTCACT	GCTCCCGCAT	AGCGAGATCCAAGT	TCACTGTG	CCCTGTGCCCGAC	540
577	AT	GGGGTCACT	GCTCCCGCAT	AGCGAGATCCAAGT	TCACTGTG	CCCTGTGCCCGAC	636
541	CAG	TTCAAGGGAAA	TTCTGTGAAATAG	GTTCGTGACT	GTCTATGTTGGCGATGGCTAC	600	
637	CAG	TTCAAGGGAAA	TTCTGTGAAATAG	GTTCGTGACT	GTCTATGTTGGCGATGGCTAC	696	
601	TC	TTACCAGGGAAA	ATGAAATAGA	CAGTCAAC	CAGCATGGTGCCTTTACTTGGAACTCC	660	
697	TC	TTACCAGGGAAA	ATGAAATAGA	CAGTCAAC	CAGCATGGTGCCTTTACTTGGAACTCC	756	
661	CAC	CTCCTCTTG	CAGGAGAA	TTACACATGTTTAT	GGAGGATGCTGAAACCCCATGGGATT	720	
757	CAC	CTCCTCTTG	CAGGAGAA	TTACACATGTTTAT	GGAGGATGCTGAAACCCCATGGGATT	816	
721	GG	GGAACACA	ATTTCTGCAGAA	ACCCAGATCG	CGACGAAAGCCCTGGTCTTTATTTAA	780	
817	GG	GGAACACA	ATTTCTGCAGAA	ACCCAGATCG	CGACGAAAGCCCTGGTCTTTATTTAA	876	
781	GT	TACCAATG	CACAGGTGAA	ATGGGAATCTGTGAT	GTCTCAGCTCTCAGCCCGAC	840	
877	GT	TACCAATG	CACAGGTGAA	ATGGGAATCTGTGAT	GTCTCAGCTCTCAGCCCGAC	936	
841	GT	TGCTAC	CCAGAGGAAAG	CCCCACCTGAGCCAT	CAACCAAGCTTCCGGGGTTTGA	900	
937	GT	TGCTAC	CCAGAGGAAAG	CCCCACCTGAGCCAT	CAACCAAGCTTCCGGGGTTTGA	996	
901	TG	TGAAAGACT	GTGATAG	CAGAGGAA	GATCAAGAGAA	TCTATGGAGGCTTTAAGAC	960
997	TG	TGAAAGACT	GTGATAG	CAGAGGAA	GATCAAGAGAA	TCTATGGAGGCTTTAAGAC	1056
961	AG	GGGGCAAG	CCACCA	TGGCAGCGCT	CCCTCCAGTCC	CTGCTGACCACTCC	1020

/note="KR; Region: Kringle domain"
/db_xref="CDD:smart00130"
1017. .1730
/note="Tryp_SPC; Region: Trypsin-like serine protease"
/db_xref="CDD:smart00020"

ORIGIN

Query Match 99.9%; Score 1681.4; DB 9; Length 2251;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1682; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ATGTTTCCAGGAGTCTGATCTCCATGTTCTGCTGTTAATGGCTCTGGTGGGAAGACA 60
Db 81 ATGTTTCCAGGAGTCTGATCTCCATGTTCTGCTGTTAATGGCTCTGGTGGGAAGACA 140
Qy 61 GCCTGGGTTCTCCCTGATGCTTTTATGGAAGCCTGGAACCCAGACTGGACCCCTGAC 120
Db 141 GCCTGGGTTCTCCCTGATGCTTTTATGGAAGCCTGGAACCCAGACTGGACCCCTGAC 200
Qy 121 CAGTATGATTACAGTACAGGAGTATAATCAGGAAGAGAACACCAAGTAGCACACTTACC 180
Db 201 CAGTATGATTACAGTACAGGAGTATAATCAGGAAGAGAACACCAAGTAGCACACTTACC 260
Qy 181 CATGCTGAGAAATCTGACTGGTACTACACTGAGGACCAAGTGAATGCCAGCCCAAC 240
Db 261 CACGCTGAGAAATCTGACTGGTACTACACTGAGGACCAAGTGAATGCCAGCCCAAC 320
Qy 241 CCCTGTGAACACGCTGGGAGCTGCTCGTCCATGGGAGACCTTCAATCAGCTGCGCTG 300
Db 321 CCCTGTGAACACGCTGGGAGCTGCTCGTCCATGGGAGACCTTCAATCAGCTGCGCTG 380
Qy 301 GCTCCTTTCTCGGAATAAGTGTGAGAAAGTGCAGAAATGCAAAATACGTGCAAGGACAA 360
Db 381 GCTCCTTTCTCGGAATAAGTGTGAGAAAGTGCAGAAATGCAAAATACGTGCAAGGACAA 440
Qy 361 GGCCGGGGCAATCTCATATACCCAGAGTCCCTCCCTACTACCGCTGCTGTGAACAC 420
Db 441 GGCCGGGGCAATCTCATATACCCAGAGTCCCTCCCTACTACCGCTGCTGTGAACAC 500
Qy 421 CTTTACACAGTCCAGCTGCTCCCAAGTGGTCTCTGTATGAGGCAAAACCCCTGCGAG 480
Db 501 CTTTACACAGTCCAGCTGCTCCCAAGTGGTCTCTGTATGAGGCAAAACCCCTGCGAG 560
Qy 481 AATGGGCTACTGCTCCCGCATAGGGAGATCCAGTTCACCTGTGCTGCTGCTCCGAC 540
Db 561 AATGGGCTACTGCTCCCGCATAGGGAGATCCAGTTCACCTGTGCTGCTGCTCCGAC 620
Qy 541 CAGTTCAAGGGGAAATCTGTGAATAAGTTCCTGATGACTGCTATGTTGGCGATGGCTAC 600
Db 621 CAGTTCAAGGGGAAATCTGTGAATAAGTTCCTGATGACTGCTATGTTGGCGATGGCTAC 680
Qy 601 TCTTACCGAGGAAATGAATAGGACAGTCAACAGCATGCGTCCCTTTACTGGAATCC 660
Db 681 TCTTACCGAGGAAATGAATAGGACAGTCAACAGCATGCGTCCCTTTACTGGAATCC 740
Qy 661 CACCTCCTCTTTCAGGAGAAATCAACATGTTTATGGAGATGCTGAAACCCATGGGATT 720
Db 741 CACCTCCTCTTTCAGGAGAAATCAACATGTTTATGGAGATGCTGAAACCCATGGGATT 800
Qy 721 GGGGAACAAATTTCTGAGAAACCCAGATCGGACGAAAGCCCTGGTCTTTATTTAAA 780
Db 801 GGGGAACAAATTTCTGAGAAACCCAGATCGGACGAAAGCCCTGGTCTTTATTTAAA 860
Qy 781 GTTACCAATGACAAAGTGAATGGAAATGCTGATGCTCAGGCTGCTCAGCCAGGAC 840
Db 861 GTTACCAATGACAAAGTGAATGGAAATGCTGATGCTCAGGCTGCTCAGCCAGGAC 920
Qy 841 GTTGCTTACCCAGAGGAAAGCCCACTGAGCATCAACCAAGCTTCGGGGTTGACTCC 900
Db 921 GTTGCTTACCCAGAGGAAAGCCCACTGAGCATCAACCAAGCTTCGGGGTTGACTCC 980
Qy 901 TGTGAAAGACTGAGATAGCAGAGGAGAGATCAAGAAATCTATGAGGCTTTAAGAGC 960

Db 981 TGTGAAAGACTGAGATAGCAGAGAGGAGAAATCAAGAGAAATCTATGGAGGCTTTAAGAGC 1040
Qy 961 ACGGCGGCAAGCACCCATGSCAGCGTCCCTCAGTCTCTGCTGCTCTGACCATCTCC 1020
Db 1041 ACGGCGGCAAGCACCCATGSCAGCGTCCCTCAGTCTCTGCTGCTCTGACCATCTCC 1100
Qy 1021 ATGCCCCAGGCGCACTTCTGTGGTGGGCGCTGATCCACCCCTGCTGGGTGCTCACTGCT 1080
Db 1101 ATGCCCCAGGCGCACTTCTGTGGTGGGCGCTGATCCACCCCTGCTGGGTGCTCACTGCT 1160
Qy 1081 GCCCACTGCACCGACATAAAAAACCCAGACATCTAAAGGTGGTCTAGGGAGCCAGGACCTG 1140
Db 1161 GCCCACTGCACCGACATAAAAAACCCAGACATCTAAAGGTGGTCTAGGGAGCCAGGACCTG 1220
Qy 1141 AAGAAGAGAAATTTCAATGACGACAGCTTTTAGGGTGGAGAGATATTCAGTACAGCCAC 1200
Db 1221 AAGAAGAGAAATTTCAATGACGACAGCTTTTAGGGTGGAGAGATATTCAGTACAGCCAC 1280
Qy 1201 TACATGAAAGAGATGAGATTCCCAATGATATTGCTCAAGTCTCAAGTCTAAAGTCTCAAGT 1260
Db 1281 TACATGAAAGAGATGAGATTCCCAATGATATTGCTCAAGTCTCAAGTCTAAAGTCTCAAGT 1340
Qy 1261 GATGCTACTGCTCTAGAAATCCAAATACCTGAAGACTGTGCTGCTGCTGCTGCTGCTGCTGCT 1320
Db 1341 GATGCTACTGCTCTAGAAATCCAAATACCTGAAGACTGTGCTGCTGCTGCTGCTGCTGCTGCT 1400
Qy 1321 TTTCCCTCTGGGAGTGAAGTGCACATCTCTGCTGGGTGGTGTACAGAAACAGGAAAGGG 1380
Db 1401 TTTCCCTCTGGGAGTGAAGTGCACATCTCTGCTGGGTGGTGTACAGAAACAGGAAAGGG 1460
Qy 1381 TCCGCCAGCTCTCGGATGCCAAAGTCAAGTGAATGCCAACACTTTGTGCAATCCCGC 1440
Db 1461 TCCGCCAGCTCTCGGATGCCAAAGTCAAGTGAATGCCAACACTTTGTGCAATCCCGC 1520
Qy 1441 CAACTCTATGACCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1500
Db 1521 CAACTCTATGACCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1580
Qy 1501 GGGGAAGACACTGCGAGGCTGACTCTGAGAGCCCTGACTGCTGAGAGAGAGAGAGAGAGAGAGAG 1560
Db 1581 GGGGAAGACACTGCGAGGCTGACTCTGAGAGCCCTGACTGCTGAGAGAGAGAGAGAGAGAGAGAG 1640
Qy 1561 TACTACGCTATGGGATAGTGAAGTGGGCTGAGTGGGCTGAGTGGGAGAGAGAGAGAGAGAGAGAG 1620
Db 1641 TACTACGCTATGGGATAGTGAAGTGGGCTGAGTGGGCTGAGTGGGAGAGAGAGAGAGAGAGAGAG 1700
Qy 1621 ACCCAAGTTACAAATTCCTGAATTCAGATCAAAAGCCACCATCAAAAGTGAAGTGGCTTC 1680
Db 1701 ACCCAAGTTACAAATTCCTGAATTCAGATCAAAAGCCACCATCAAAAGTGAAGTGGCTTC 1760
Qy 1681 TAA 1683
Db 1761 TAA 1763

RESULT 7
AX383955
LOCUS AX383955
DEFINITION Sequence 2 from Patent EP1182258.
ACCESSION AX383955
VERSION AX383955.1 GI:19577526
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1
REFERENCE
AUTHORS Roenisch, J., Stoeck, H.A., Feussner, A., Lang, W., Weimer, T.,
Becker, M., Nerlich, C. and Muth-Naumann, G.
TITLE Mutants of the factor vii activating protease and methods for their
detection
JOURNAL Patent: EP 1182258-A 2 27-FEB-2002;
Aventis Behring GmbH (DE)

901	TGTGGAAGAC	CTGAGATAGCAGAGAGGAAAGATCAAGAGAAATCTATGGAGGCTTTAAAGGC	960
961	ACGGCGGGCAAGC	ACCCATGCGCAGCGCTCCCTCCAGTCCCTCGCTGCCTCTGACCACTCTCC	1020
961	ACGGCGGGCAAGC	ACCCATGCGCAGCGCTCCCTCCAGTCCCTCGCTGCCTCTGACCACTCTCC	1020
1021	ATGCCCCAGGGCCACT	CTCTGTGGTGGGGCGCTGATCCACCCCTGCTGGGTGCTCACTGCT	1080
1021	ATGCCCCAGGGCCACT	CTCTGTGGTGGGGCGCTGATCCACCCCTGCTGGGTGCTCACTGCT	1080
1081	GCCCACTGCACCGACAT	AAAAACACAGACATCTAAAGGTGGTGTAGGGGACACGAGACCTTG	1140
1081	GCCCACTGCACCGACAT	AAAAACACAGACATCTAAAGGTGGTGTAGGGGACACGAGACCTTG	1140
1141	AGAAAGAAGAAATTT	TCATGAGCAGAGCTTTAGGGTGGAGAATATTTCAAGTACAGCCAC	1200
1141	AGAAAGAAGAAATTT	TCATGAGCAGAGCTTTAGGGTGGAGAATATTTCAAGTACAGCCAC	1200
1201	TACAATGAAAGAGATG	AGATTTCCCCACAAATGATTTGCAATTTGCTCAAGTTAAAGCCAGTG	1260
1201	TACAATGAAAGAGATG	AGATTTCCCCACAAATGATTTGCAATTTGCTCAAGTTAAAGCCAGTG	1260
1261	GATGTCTACTGTCTCT	AGAAATCCAAATACGTGAAGACTGTGTGCTTGCTGATGGGTCC	1320
1261	GATGTCTACTGTCTCT	AGAAATACGTGAAGACTGTGTGCTTGCTGATGGGTCC	1320
1321	TTTCCCTCTGGAGTGA	GTGCCACATCTCTGGCTGGGGTGTTCACAGAAACAGGAAAGGG	1380
1321	TTTCCCTCTGGAGTGA	GTGCCACATCTCTGGCTGGGGTGTTCACAGAAACAGGAAAGGG	1380
1381	TCCCGCCAGCTCCT	GGATGCCAAAGTCAAGCTGATTGCCAACACTTTGTGCAACTCCCCG	1440
1381	TCCCGCCAGCTCCT	GGATGCCAAAGTCAAGCTGATTGCCAACACTTTGTGCAACTCCCCG	1440
1441	CAACTCTATGACCAAT	GTGATGACAGTATGATCTGTGCAGGAAATCTTCAGAAACCT	1500
1441	CAACTCTATGACCAAT	GTGATGACAGTATGATCTGTGCAGGAAATCTTCAGAAACCT	1500
1501	GGGCAAGACACCTGC	CCAGGGTGACTCTGGAGGCCCTCCACCTGTGAGAAAGCAGGCACC	1560
1501	GGGCAAGACACCTGC	CCAGGGTGACTCTGGAGGCCCTCCACCTGTGAGAAAGCAGGCACC	1560
1561	TACTAGCTCTATGGG	ATAGTAGTGGGCTGGAGTGTGAGAAAGGCGGGGTCTAC	1620
1561	TACTAGCTCTATGGG	ATAGTAGTGGGCTGGAGTGTGAGAAAGGCGGGGTCTAC	1620
1621	ACCCAGTTACCAAAAT	TCCTGAATGGATCAAAAGCCACCATCAAAAGTGAAGTGGCTTC	1680
1621	ACCCAGTTACCAAAAT	TCCTGAATGGATCAAAAGCCACCATCAAAAGTGAAGTGGCTTC	1680
1681	TAA	1683	
1681	TAA	1683	
RESULT 8			
BD177603	1693 bp DNA linear PAT 16-APR-2003		
LOCUS	Detection method with the use of factor VII activating protease		
DEFINITION	mutant and specific antibody.		
ACCESSION	BD177603		
VERSION	BD177603.1 GI:30014865		
KEYWORDS	JP 2002291486-A/2.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	Roemisch, J., Stoeck, H.A., Feusner, A., Lang, W., Weimer, T., Becker, M., Nethlich, C., and Naumann, G.M.		
TITLE	Detection method with the use of factor VII activating protease mutant and specific antibody		
JOURNAL	Patent: JP 2002291486-A 2 08-OCT-2002;		

[illegible]

RESULT 10					
EC031775	LOCUS	BC031775	2119 bp	mRNA	linear
DEFINITION					Mus musculus hyaluronic acid binding protein 2, mRNA (cDNA clone
					MGC:28705 IMAGE:4242577), complete cds.
					ROD 04-OCT-2003

BC031775.1	GI:21594554
VERSION	
KEYWORDS	
MGC	
Mus musculus (house mouse)	
ORGANISM	
Mus musculus	
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;	
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus;	
1 (bases 1 to 2119)	
REFERENCE	
STRANSBERG, R.L., FEINGOLD, E.A., GROUSE, L.H., DERSE, J.G.	
AUTHORS	

Atsushi, K.B., Corliss, F.S., Wegner, L., Suenmer, C.M., Schuler, S.D.,
 Atschul, S.F., Zebberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,
 Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,
 Diatchenko, L., Marasina, K., Farmer, A.A., Rubin, G.M., Hong, L.,
 Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,
 Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S.,
 Carninci, P., Prange, C., Raha, S.S., Loquellano, N.Y., Peters, G.J.,

McKernan, A. J., Mullany, S. O., Bosak, S. A., McWaters, P. O.,
O'Kearney, K. J., Malek, A. A., Gunaratne, P. H., Richards, S.,
Worley, K. C., Hale, S., Garcia, A. M., Gay, L. J., Hulyk, S. W.,
Villalón, D. K., Muzny, D. M., Sodergren, E. J., Lu, X., Gibbs, R. A.,
Fahey, J., Helton, E., Ketterman, M., Madan, A., Rodrigues, S.,
Sanchez, A., Whiting, M., Madan, A., Young, A. C., Shevchenko, Y.,
Bouffard, G. G., Blakesley, R. W., Touchman, J. W., Green, B. D.,
Dickson, M. C., Rodriguez, A. C., Grimwood, J., Schmutz, J., Myers, R. M.,
Butterfield, Y. S., Krzywinski, M. I., Skalska, U., Smalley, D. E.,
Schnerber, A., Schein, J. E., Jones, S. J., and Matra, M. A.

JOURNAL TITLE
Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

PUBMED 12477932
REFERENCE 2 (bases 1 to 2119)
AUTHORS Strausberg, R.

Submitted (06-JUN-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

REMARK	COMMENT
NIH-MGC Project URL: http://mgc.nci.nih.gov	
Contact: MGC help desk	
Email: cgaps-r@mail.nih.gov	
Tissue Procurement: Jeffrey E. Green, M.D.	
cDNA Library Preparation: Life Technologies, Inc.	
cDNA library Arraved by: The I.M.A.G.E. Consortium	
	(LLNL)

QY 1541 CCTGTGAGAGGAGCGCACCTACTACTGCTATGGGATAGTGGGCGCTGGAGTGTG 1600
Db 1470 CCTGTGAGAGGATGGAAGCTTACTAGCTTACGGGATTTAAGCTGGGCGCAGGAATGTG 1529
QY 1601 GGAAGAGGCGAGGGGTCTACACCCAGTTACCAAAATTCCTGAATTGGATCAAGCCACCA 1660
Db 1530 GGAAGAGGCGAGGCTTACACTCAAGTCAAGTCTCTGAATTGGAATAAGACCA 1589
QY 1661 TCAAAAGTGAAGTGGCTTCTAA 1683
Db 1590 TGCACAGGAGGCTGGCCTCTGA 1612

RESULT 11
AL390197 188349 bp DNA linear PRI 29-AUG-2001
LOCUS Human DNA sequence from clone RP11-190F19 on chromosome 10,
DEFINITION complete sequence.
ACCESSION AL390197
VERSION AL390197.19 GI:15387800
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 188349)
AUTHORS Tracey,A.
DIRECT SUBMISSION
SUBMITTED (29-AUG-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
REQUESTS: clonesrequest@sanger.ac.uk
On Aug 31, 2001 this sequence version replaced gi:14596344.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em:, EMBL; Sw:,
SWISSPROT; Tr:, TrEMBL; Wp:, WORMPEP; Information on the WORMPEP
database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 10, constructed by the Sanger Centre Chromosome 10
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr10
RP11-190F19 is from the library RPI-11.1 constructed by the group
of Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pBACE3.6
This sequence is the entire insert of clone RP11-190F19 The true
right end of clone RP11-481H2 is at 111825 in this sequence.

FEATURES
Location/Qualifiers
source
1. 188349
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosomes="10"
/clone="RP11-190F19"
/clone_lib="RPI-11.1"
22142..22187
/notes="Sequence from overlapping clone RP11-481H2
(AL133482). Assembly confirmed by restriction digest."

misc_feature
ORIGIN

QY 461 GCAGCCCAAAACCCCTGCCAGATGGGCTACCTGCTCCCGGCATAGCGGAGATCCAAAGT 520
Db 390 GCAGCCCAAAACCCCTGCCAGATGGGCTACCTGCTCCCGGCATAGCGGAGATCCAAAGT 449
QY 521 TCACCTGTGCTGTCCGACGAGTTCAAGGGGAATTCGTGAATAGTGTCTGATGACT 580
Db 450 TTACTGTGCTGTCCGACGAGTTCAAGGGGAATTCGTGAATAGTGTCCGACGACT 509
QY 581 GCTATGTTGGGATGGCTACTCTTACCGAGGGAATCAATAGGACAGTCAACACGATG 640
Db 510 GTTATGTCGGTGTGATGGCTACTCTTACCGAGGCAAGTGAAGACAGTCAACAGAAC 569
QY 641 CGTGCCTTTACTGGAACTCCCACTCTCTTGCAGGAGAAATTAACAATGTTTATGAGG 700
Db 570 CATGCTTTACTGGAACTCCCACTCTCTTGCAGGAGACTTATAACATGTTTATGAGG 629
QY 701 ATGCTGAAACCCATGGGATGGGGAACACAATTTCTGCAGAAACCCAGATCGGACGAAA 760
Db 630 ATGCAGAGACCCACGGATCGCAGAGCACAACTTCTGCAGAAACCCAGATCGGACGACA 689
QY 761 AGCCCTGTGCTTTTAAAGTATCAATGACAAGTGAATGGGAATPACTGTGATGCT 820
Db 690 AACCTGTGTTTCTGCAAGTGAACAGTGAAGGTGAATGGGAATPACTGTGATGCTCA 749
QY 821 CAGCCTGTCTAGCCAGGACGTTGCTCTACCCAGAGGAAAGCCCACTGAGCCATCAACCA 880
Db 750 CAGTCTGTCCAGTCTGACACCCCTTACCCAGTGAAGGCTTCTGGAGCTGTGATGG 809
QY 881 AGCTTCCGGGTTGACTCTCTGTGAAGAAGCTGAGATAGCAGAGGAAGATCAAGAGAA 940
Db 810 AGCTGCCAGGTTTGAAGTCTCTCGGGAAGACGAGGTAGCTGAACACGACGCTCAAGCGTA 869
QY 941 TCTATGAGGCTTTAAGAGCAGCGGCGGCAAGCACCCATGGCAGCGCTCCCTCCAGTCT 1000
Db 870 TCTACGGGGCTTTAAGAGCAGCAGCAGGACGACCCGCTGGCAGTGTCTCTGAGACT 929
QY 1001 CGCTGCTCTGACCTTCCATCTCCCAAGGCGCACTTCTGTGGTGGGCGCTGATCCACC 1060
Db 930 CACTGCTGTGACCACTCCATCTCCCAAGGCGCACTTCTGTGGGCGCGCTGATCCACC 989
QY 1061 CTGTGGTGGTCTACTGTGCGCCACTCCGACATAAACACGACATCTTAAGGTGG 1120
Db 990 CTGTGGTGGTCTACTGTGCGCCACTGTACGACATAAACACGACATCTTAAGGTGG 1049
QY 1121 TGCTAGGGAACGAGCACTTGAAGAAGAAATTTTCATGACGAGAGCTTTAGGGTGGAGA 1180
Db 1050 TACTAGGGAACGAGCACTTGAAGAAGAAATCTCCATGAACAGACCTTCAGGGTGGAAA 1109
QY 1181 AGATATTCAGTACAGCCACTCAATGAAGATGAGATTTCCCAAGATGATTTGCAT 1240
Db 1110 AAATACTGAAGTACAGTCAATGAATGAAGAGATGAGATTTCCCAAGATGATTTGCTT 1169
QY 1241 TGCTCAAGTTAAAGCCAGTGGTGTCTACTGTGCTCTAGAAATCCAAATACGTAAGACTG 1300
Db 1170 TGCTCAAGTTAAAGCCAGTGGTGTCTACTGTGCTCTAGAAATCCAGATATGTAAGACTG 1229
QY 1301 TGCTGTGCTGATGGTCTTCCCTCTGGGAGTGAAGTGCACATCTCTGGCTGGGCTG 1360
Db 1230 TATGTTTGCCCGACGACCCCTTCCCTCTGGAACAGTGAAGTGCACATCTCTGGCTGGGCTG 1289
QY 1361 TTACAGAAACAGAAAGGTCCTGCTGATGCCAAAGTCAAGCTGATTTGCCA 1420
Db 1290 TTACAGAAACAGGGAAGGTCCTGCTGATGCCAAAGTCAAGCTGATTTGCCA 1349
QY 1421 ACATTTGTGCACTCCCGCCCACTCTATGAACAATGATGATGATGATGATGATGATG 1480
Db 1350 ACCCTTTGTGCACTCCCGCCCACTCTATGAACAATGATGATGATGATGATGATGATG 1409
QY 1481 CAGGAATCTTCAGAACTCTGGCAGACACTGCGAGGTCACCTCTGGAGGCCCCCTGA 1540
Db 1410 CGGGGAACCTTCAGAAAGCCCGGATCAGACCTTCCAGGGTGAATTCGGGGGGCCCTCTAA 1469


```

* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 196130: contig of 196130 bp in length
* 196131 196230: gap of unknown length
* 196231 224208: contig of 27978 bp in length.
FEATURES
    source
        1..224208
            /organism="Rattus norvegicus"
            /mol_type="genomic DNA"
            /db_xref="taxon:10116"
            /clone="CH230-131J10"
            2668..3129
                /note="clone boundary"
                clone_end:Sp6
                site:EcoRI
            end sequence: BH346758"
            200382..201926
                /note="wgs_contig"

    misc_feature
        200382..201926
            /note="wgs_contig"

    ORIGIN
Query Match          9.6%; Score 161; DB 2; Length 224208;
Best Local Similarity 76.7%; Pred. No. 2.4e-36;
Matches 197; Conservative 0; Mismatches 60; Indels 0; Gaps 0;
QY 838 GACGTTGCTACCCAGAGGAAGCCCACTGAGCCATCAACCAAGCTTCGGGGTTTGAC 897
Db 71048 GATGCTGTATACCCATTGGGAGCCTTCAGGAACCTGTGATGGAGCTGCCAGGATTCGAC 70989
QY 898 TCCTGTGGAAGACTGAGATAGCAGAGAGAGGAAGATCAAGAGAATCTATGGAGGCTTTAAG 957
Db 70988 TCCTGCGGGAAGACAGAGATGACTCAACACGCGGTCAAGCGCATCTACGGGGCTTTAAG 70929
QY 958 AGCAGCGGGGCAAGCACCACCGGCGTCCCTCCAGTCTCGCTGCTGACCATC 1017
Db 70928 AGCAGCGGGGCAAGCACCCTGGGAGGTGCCCTGCGAGCTTATTCGCGCTGACCACC 70869
QY 1018 TCCATGCCCCAGGGCCACTTCTGTGGTGGGGCGGTGATCCACCCCTGCTGGGTGCTCACT 1077
Db 70868 TCCATGCCCCAAGGCCACTTCTGTGGGGGTTCCCTGATTCACCCCTTGTGGTGTCTCACT 70809
QY 1078 GCTGCCCACTGCACCGA 1094
Db 70808 GCAGCCCACTGTACCGA 70792

Search completed: May 25, 2004, 09:00:30
Job time : 4486.5 secs

```

OM nucleic - nucleic search, using sw model

Run on: May 25, 2004, 04:19:43 ; Search time 471 Seconds
(without alignments)
15179.867 Million cell updates/sec

Title: US-09-912-559-1
Perfect score: 1683
Sequence: 1 atgtttgcaggatgtctga.....aaagtgaagtggttcttaa 1683

Scoring table: IDENTITY_NUC
Gapop 10_0 , Gapext 1.0

Searched: 337863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N Geneseq 29Jan04: *
1: Geneseqm1980s: *
2: Geneseqm1990s: *
3: Geneseqm2000s: *
4: Geneseqm2001as: *
5: Geneseqm2001bs: *
6: Geneseqm2002s: *
7: Geneseqm2003as: *
8: Geneseqm2003bs: *
9: Geneseqm2003cs: *
10: Geneseqm2004s: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1683	100.0	1683	6	AAL45696 Human blo
2	1683	100.0	3008	6	ABN95170 Gene #166
3	1683	100.0	3008	7	ACC51204 Human Plk
4	1679.8	99.8	1683	6	AAL45697 Human blo
5	1678.2	99.7	3623	3	AACT6693 Human ORF
6	526.4	31.3	617	3	AAA44763 Human sec
7	368.4	21.9	428	8	ACH21030 Human adu
8	242.6	14.4	397	6	AB199281 Mouse isc
9	120.2	7.1	451	4	ABA58323 Human foe
10	120.2	7.1	451	4	AAI37947 Probe #66
11	120.2	7.1	451	4	AAK32096 Human bon
12	120.2	7.1	451	4	AAK06420 Human bra
13	120.2	7.1	451	4	ABS31791 Human liv
14	120.2	7.1	451	6	ABS06863 Human gen
15	117.6	7.0	2036	2	AAQ63951 Hepatocyt
16	117.6	7.0	2036	6	ABL65242 Lung canc
17	117.6	7.0	2036	6	ABN95114 Gene #161
18	117	7.0	117	4	ABA70919 Human foe
19	117	7.0	117	4	AAI51106 Probe #19
20	117	7.0	117	4	AAK45153 Human bon
21	117	7.0	117	4	AAK19188 Human bra
22	117	7.0	117	4	ABS44824 Human liv
23	117	7.0	117	6	ABS19403 Human gen

24	102.6	6.1	2145	9	ADSE06939	Ade06939	Novel cod
25	96.2	5.7	970	2	AAQ63945	Gene comp	
26	93.2	5.5	2299	7	ABX17688	DNA encod	
27	90.6	5.4	2512	2	AAQ12867	JMI-229 c	
28	89	5.3	2445	9	ADBS2711	Primary r	
29	87.6	5.2	2252	2	AAQ05673	v-PA alph	
30	86.8	5.2	1323	1	AAK70992	Modified	
31	85.2	5.1	1422	1	AAK70991	Modified	
32	84.8	5.0	2137	2	AAQ05675	v-PA_beta	
33	79.6	4.7	2257	2	AAQ05674	Chicken u	
34	78.8	4.7	919	1	AAK90048	Mutated c	
35	78.6	4.7	1689	1	AAK91217	AAK91217	
36	78	4.6	1620	2	AAQ00543	Vampire b	
37	77	4.6	2037	2	AAQ12071	T-PA vari	
38	77	4.6	2100	2	AAQ12074	T-PA vari	
39	75.4	4.5	1047	2	AAQ40658	TPA-2 CDN	
40	75.4	4.5	1065	2	AAK7456	mt-PA6-E	
41	75.4	4.5	1065	2	AAK7456	mt-PA6-E	
42	75.4	4.5	1068	2	AAQ53322	Human tru	
43	75.4	4.5	1068	2	AAQ53320	Human tru	
44	75.4	4.5	1068	2	AAQ40683	TPA-5 CDN	
45	75.4	4.5	1068	2	AAQ40657	TPA-1 CDN	

ALIGNMENTS

RESULT 1
AAL45696
ID AAL45696 standard; DNA; 1683 BP.

XX AAL45696;
AC AAL45696;
DT 13-JUN-2002 (first entry)
XX Human blood coagulation factor VII activating protease DNA.
DE Human; blood coagulation factor VII activating protease; FSAP;
KW single-chain plasminogen activator; bleeding disorder; haematological;
KW haemostatic; gene; ds.
XX Homo sapiens.
XX Key Location/Qualifiers
FT CDS 1..1683
FT /tag= a
FT /product= "FSAP"
XX
PN EP1182258-A1.
XX 27-FEB-2002.
XX 05-JUL-2001; 2001EP-00115691.
XX 26-JUL-2000; 2000DE-01036641.
XX 10-OCT-2000; 2000DE-01050040.
XX 21-OCT-2000; 2000DE-01052319.
XX 12-APR-2001; 2001DE-01018706.
XX (AVET) AVENTIS BEHRING GMBH.
XX Roemisch J, Stoehr H, Feussner A, Lang W, Weiner T, Becker M;
XX Nerlich C, Muth-Naumann G;
XX WPI; 2002-270939/32.
XX P-PSDB; AAO17144.
XX New nucleic acid encoding mutant factor 7 activating protease, useful for
XX diagnosis, treatment and prevention of coagulation disorders, also
XX related protein and antibodies.
XX Claim 2; Page 15-16; 27pp; German.

CC The present invention relates to a mutant of the DNA sequence encoding
CC the protease (FSAP) that activates blood coagulation factor VII (FVII)
CC and single-chain plasminogen activator, where at least one of the base
CC changes G to C at nucleotide 1177 and G to A at nucleotide 1601 is
CC present. The mutant sequences can be used in the treatment and prevention
CC of bleeding disorders associated with inherited or acquired defects of
CC blood coagulation factors V, VII, IX, X, XI, XII, von Willebrand's
CC factor and/or with antibodies against any of these proteins. The present
CC sequence is the human FSAP coding sequence

```
Query Match      100.0%; Score 1683; DB 6; Length 1683;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1683; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

QY	1	ATGTTGCCAGATGTCGTGATCTCCATGTTCTGCTGTATGCTCTGCTGGGAAGACA	60
DB	1	ATGTTTGCAGAGTGTCTGATCTCCATCTGTTCTGCTGTATGCTCTGCTGGGAAGACA	60
QY	61	GCCTGTGGGTTCTCCCTGATGTCCTTTATTGGAAGCCTGGACCAGACTGGACCCCTGCAC	120
DB	61	GCCTGTGGGTTCTCCCTGATGTCCTTTATTGGAAGCCTGGACCAGACTGGACCCCTGCAC	120
QY	121	CAGTATGATTACAGCTACGAGGATATTAATCAGGAAGAGAACACCACTAGTAGCACATTACC	180
DB	121	CAGTATGATTACAGCTACGAGGATATTAATCAGGAAGAGAACACCACTAGTAGCACATTACC	180
QY	181	CATGCTGAGAACTCTGACTTGGTACTACACTGAGGACCAAGCTGATCCATGCGAGCCCAAC	240
DB	181	CATGCTGAGAACTCTGACTTGGTACTACACTGAGGACCAAGCTGATCCATGCGAGCCCAAC	240
QY	241	CCCTGTGAACAGGTTGGGACCTGCTCGTCCATCGGAGCACCTTCACATGCAGCTGSCCTG	300
DB	241	CCCTGTGAACAGGTTGGGACCTGCTCGTCCATCGGAGCACCTTCACATGCAGCTGSCCTG	300
QY	301	GCTCCTTTCTCTGGGAATAAGTGTGAGAAAGTGCAAAATACGTGCAAGGACAAACCCATGT	360
DB	301	GCTCCTTTCTCTGGGAATAAGTGTGAGAAAGTGCAAAATACGTGCAAGGACAAACCCATGT	360
QY	361	GGCCGGGGCCATGTCTCATTAACCCAGAGTCTCTCCCTACTACCGCTGTGCTGTAAACAC	420
DB	361	GGCCGGGGCCATGTCTCATTAACCCAGAGTCTCTCCCTACTACCGCTGTGCTGTAAACAC	420
QY	421	CCTTACACAGGTCCTCCAGCTGCTCCCAAGTGTTCTGTATGCAAGGCCAAAACCCCTGCCAG	480
DB	421	CCTTACACAGGTCCTCCAGCTGCTCCCAAGTGTTCTGTATGCAAGGCCAAAACCCCTGCCAG	480
QY	481	AATGGGCTACCTGCTCCCGGCATTAAGCGGAGATCCAAAGTTACCTGTGCTGTCCCGAC	540
DB	481	AATGGGCTACCTGCTCCCGGCATTAAGCGGAGATCCAAAGTTACCTGTGCTGTCCCGAC	540
QY	541	CAGTTCAAGGGGAAATTCGTGGAATAGTGTCTGATGACTGCTATGTTGGCGATGCTCTAC	600
DB	541	CAGTTCAAGGGGAAATTCGTGGAATAGTGTCTGATGACTGCTATGTTGGCGATGCTCTAC	600
QY	601	TCCTTACCGGGGAAATGAATAGGACAGTCAACAGCATGCTGCCCTTACTGGAACTCC	660
DB	601	TCCTTACCGGGGAAATGAATAGGACAGTCAACAGCATGCTGCCCTTACTGGAACTCC	660
QY	661	CACCTCCTCTTTCAGGAGAAATTACAAATGTTTATGAGGATGCTGAAAACCAATGGGATT	720
DB	661	CACCTCCTCTTTCAGGAGAAATTACAAATGTTTATGAGGATGCTGAAAACCAATGGGATT	720
QY	721	GGGGAAACAATTTCTGCAGAAACCCAGATGGGACGAAAGCCCTGGTGTCTTTATTAAA	780
DB	721	GGGGAAACAATTTCTGCAGAAACCCAGATGGGACGAAAGCCCTGGTGTCTTTATTAAA	780
QY	781	GTTACCAATGACAAAGTGAAATGGGAAATCTGTGATGTCTCAGCCTGCTCAGCCCGAC	840
DB	781	GTTACCAATGACAAAGTGAAATGGGAAATCTGTGATGTCTCAGCCTGCTCAGCCCGAC	840
QY	841	GTTGCCCTACCGAGGAAAGCCCACTGAGGCCATCAACCAAGCTTCCGGGTTTGTACTCC	900

XX Homo sapiens.
 OS WO200229103-A2.
 PN 11-APR-2002.
 PD 02-OCT-2001; 2001WO-US030589.
 PF 02-OCT-2000; 2000US-0237054P.
 PR (GENE-) GENE LOGIC INC.
 PA
 PP
 PI Horne D, Alvares C, Peres-Da-Silva S, Vockley JG;
 DR WPI; 2002-426119/45.
 XX
 XX Diagnosing and detecting the progression of liver cancer, hepatocellular
 PT carcinoma or metastatic liver tumor in a patient, involves detecting the
 PT level of expression of two or more genes in a liver tissue sample.
 XX
 PS Claim 1; SEQ ID NO 1668; 298pp; English.
 XX
 CC The invention relates to a novel method for diagnosing and detecting the
 CC progression of liver cancer, hepatocellular carcinoma or metastatic liver
 CC tumor in a patient, and differentiating metastatic liver cancer from
 CC hepatocellular carcinoma in a patient, involving detecting the level of
 CC expression of two or more genes represented in ABN93503-ABN97455 in a
 CC tissue sample. The method of the invention has hepatocytic, and
 CC cytosolic activity. The method is useful for diagnosing and detecting
 CC the progression of liver cancer, hepatocellular carcinoma and metastatic
 CC liver carcinoma in a patient. The method is useful for identifying
 CC expression profiles which serve as useful diagnostic markers as well as
 CC markers that can be used to monitor disease states, disease progression,
 CC drug toxicity, drug efficacy and drug metabolism. Note: The sequence data
 CC for this patent did not form part of the printed specification, but was
 CC obtained in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 3008 BP; 791 A; 837 C; 703 G; 677 T; 0 U; 0 Other;

Query Match 100.08; Score 1693; DB 6; Length 3008;
 Best Local Similarity 100.08; Pred. No. 0;
 Matches 1693; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGTTTGCAGGATGCTGATCTCCATGTTCTGCTGTTAATGGCTCTGGTGGGAAGACA 60
 DB 97 ATGTTGCCAGGATGCTGATCTCCATGTTCTGCTGTTAATGGCTCTGGTGGGAAGACA 156
 QY 61 GCCTGTGGGTTCTCCCTGATGCTTTTATTGGAAGCCTGGACCCAGACTGGACCCCTGAC 120
 DB 157 GCCTGTGGGTTCTCCCTGATGCTTTTATTGGAAGCCTGGACCCAGACTGGACCCCTGAC 216
 QY 121 CAGTATGATTACAGCTACGAGATTATTAATCAGGAAGAGAACACCAGTAGCACACTTACC 180
 DB 217 CAGTATGATTACAGCTACGAGATTATTAATCAGGAAGAGAACACCAGTAGCACACTTACC 276
 QY 181 CATGCTGAGATTCCTGATGTTACTACTGAGGACCAAGTGATCCATGCGCACCCCAAC 240
 DB 277 CATGCTGAGATTCCTGATGTTACTACTGAGGACCAAGTGATCCATGCGCACCCCAAC 336
 QY 241 CCTCTGGAACAGGTTGGGAGCTGCTGCTGCTCATGAGGACCTTTCACATGAGCTGCTG 300
 DB 337 CCTCTGGAACAGGTTGGGAGCTGCTGCTGCTCATGAGGACCTTTCACATGAGCTGCTG 396
 QY 301 GCTCCTTTCTCTGGGAATAGTGTGAGAAAGTCAAAATACGTGCAAGGACCAACCCATGT 360
 DB 397 GCTCCTTTCTCTGGGAATAGTGTGAGAAAGTCAAAATACGTGCAAGGACCAACCCATGT 456
 QY 361 GGCCGGGGCCATGCTCTATTACCCAGAGTCTCTCCTACTACCGTGTGCTGTGTAACAC 420
 DB 457 GGCCGGGGCCATGCTCTATTACCCAGAGTCTCTCCTACTACCGTGTGCTGTGTAACAC 516

421 CTTTACACAGGTCCTCCAGCTGCTCCCAAGTGGTTCTGTATGACAGGCCAAACCCCTGCCAG 480
 DB CTTTACACAGGTCCTCCAGCTGCTCCCAAGTGGTTCTGTATGACAGGCCAAACCCCTGCCAG 576
 QY 481 AATGGGCTTACCTGCTCCCGGCATAGCGGAGATCCAAAGTTCACCTGTGCTGCTGCCGAC 540
 DB 577 AATGGGCTTACCTGCTCCCGGCATAGCGGAGATCCAAAGTTCACCTGTGCTGCTGCCGAC 636
 QY 541 CAGTTCAAGGGGAAATTTCTGTGAAATAGGTTCTGTATGATGCTATGTTGGCGATGGCTAC 600
 DB 637 CAGTTCAAGGGGAAATTTCTGTGAAATAGGTTCTGTATGATGCTATGTTGGCGATGGCTAC 696
 QY 601 TCTTACCGAGGGGAAATGATAGGACAGTCAACACAGCATGCGTGGCTTTTACTGGAACTCC 660
 DB 697 TCTTACCGAGGGGAAATGATAGGACAGTCAACACAGCATGCGTGGCTTTTACTGGAACTCC 756
 QY 661 CACCTCCTCTTCCAGGAGAAATTAACAATGTTTATGAGGATGCTCAAAACCCATGGAT 720
 DB 757 CACCTCCTCTTCCAGGAGAAATTAACAATGTTTATGAGGATGCTCAAAACCCATGGAT 816
 QY 721 GGGGAACAATTTCTGCAGAAACCCAGATGGGAGGAAAGCCCTGTGCTTTTATAA 780
 DB 817 GGGGAACAATTTCTGCAGAAACCCAGATGGGAGGAAAGCCCTGTGCTTTTATAA 876
 QY 781 GTTACCAATGACAAGGTGAATGGGAATCTGTGATGTTCTCAGCCTGTCTCAGCCCGAGAC 840
 DB 877 GTTACCAATGACAAGGTGAATGGGAATCTGTGATGTTCTCAGCCTGTCTCAGCCCGAGAC 936
 QY 841 GTTGCTTCCAGAGGAAAGCCCTAGTCCATCAACCAAGCTTCCGGGTTTGAATCC 900
 DB 937 GTTGCTTCCAGAGGAAAGCCCTAGTCCATCAACCAAGCTTCCGGGTTTGAATCC 996
 QY 901 TGTGGAACACTGAGATAGCAGAGAGGAAAGATCAAGAGAAATCTATGGAGCTTTAAGAGC 960
 DB 997 TGTGGAACACTGAGATAGCAGAGAGGAAAGATCAAGAGAAATCTATGGAGCTTTAAGAGC 1056
 QY 961 AGCGGGGCAAGCACCATGCGGAGGCTCCCTCCAGTCCCTGCTGCTGACCAATCTCC 1020
 DB 1057 AGCGGGGCAAGCACCATGCGGAGGCTCCCTCCAGTCCCTGCTGCTGACCAATCTCC 1116
 QY 1021 ATGCCCCAGGGCCACTTCTGTGTTGGGGCGCTGATCCACCCCTGTGCTGCTGCTGCT 1080
 DB 1117 ATGCCCCAGGGCCACTTCTGTGTTGGGGCGCTGATCCACCCCTGTGCTGCTGCTGCT 1176
 QY 1081 GCCCACTGCACGACATAAACCAGACATCTAAAGGTGGTGTCTAGGGGACAGGACCTG 1140
 DB 1177 GCCCACTGCACGACATAAACCAGACATCTAAAGGTGGTGTCTAGGGGACAGGACCTG 1236
 QY 1141 AAGAAAGAAATTTTCATGAGCAGAGCTTTAGGGTGGAGAGATATCAAGTACAGCCAC 1200
 DB 1237 AAGAAAGAAATTTTCATGAGCAGAGCTTTAGGGTGGAGAGATATCAAGTACAGCCAC 1296
 QY 1201 TACAATGAAAGAGATGAGATTTCCCAACAATGATATTGCAATTGCTCAAGTTAAAGCCAGTG 1260
 DB 1297 TACAATGAAAGAGATGAGATTTCCCAACAATGATATTGCAATTGCTCAAGTTAAAGCCAGTG 1356
 QY 1261 GATGTCCTGCTGCTAGAAATCCAAATACGTGAAGACTGTGCTGCTGCTGCTGCTGCTGCT 1320
 DB 1357 GATGTCCTGCTGCTAGAAATCCAAATACGTGAAGACTGTGCTGCTGCTGCTGCTGCTGCT 1416
 QY 1321 TTTTCCCTCTGGGAGTGGAGTCCACATCTCTGCTGGGGTGTGTACAGAAACAGGAAAGGG 1380
 DB 1417 TTTTCCCTCTGGGAGTGGAGTCCACATCTCTGCTGGGGTGTGTACAGAAACAGGAAAGGG 1476
 QY 1381 TCCCGCCAGCTCCTGATGCCAAAGTCAAGCTGATTTGCCAAACACTTTTGTGCAACTCCCGC 1440
 DB 1477 TCCCGCCAGCTCCTGATGCCAAAGTCAAGCTGATTTGCCAAACACTTTTGTGCAACTCCCGC 1536
 QY 1441 CAACTCTATGACCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1500
 DB 1537 CAACTCTATGACCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1596
 QY 1501 GGGCAAGACACCTGCCAGGGTGAATCTGAGAGCCCTGCTGCTGAGAGGACGGCACC 1560


```
QY 1021 ATGCCCCAGGCCACTTCTGTGTGGGCGCTGATCCACCCCTCTGGTGCTCACTGCT 1080
Db 1117 ATGCCCCAGGCCACTTCTGTGTGGGCGCTGATCCACCCCTCTGGTGCTCACTGCT 1176
QY 1081 GCCCACTCCAGCATATAAAACAGACATCTAAAGTGTGCTAGGGACCGACCTG 1140
Db 1177 GCCCACTCCAGCATATAAAACAGACATCTAAAGTGTGCTAGGGACCGACCTG 1236
QY 1141 AAGAAAGAAATTTTCATGACGAGCTTTAGGGTGGAGAGATATTCAAGTACAGCCAC 1200
Db 1237 AAGAAAGAAATTTTCATGACGAGCTTTAGGGTGGAGAGATATTCAAGTACAGCCAC 1296
QY 1201 TACAATGAAGAGATGAGATCCCAACAATGATATTGCAATTGCTCAGTTAAAGCCAGTG 1260
Db 1297 TACAATGAAGAGATGAGATCCCAACAATGATATTGCAATTGCTCAGTTAAAGCCAGTG 1356
QY 1261 GATGTCTACTGTCTTGAATCCAAATACGTAAGACTGTGTGCTTGCCTGATGGTCC 1320
Db 1357 GATGTCTACTGTCTTGAATCCAAATACGTAAGACTGTGTGCTTGCCTGATGGTCC 1416
QY 1321 TTTCCTCTGGAGTGAAGTGCACATCTCTGGCTGGGTGTTTACAGAAACAGGAAAGGG 1380
Db 1417 TTTCCTCTGGAGTGAAGTGCACATCTCTGGCTGGGTGTTTACAGAAACAGGAAAGGG 1476
QY 1381 TCCGCCAGCTCTCGATGCCAAAGTCAAGCTGATTGCCAACATTTGTGCAACTCCCGC 1440
Db 1477 TCCGCCAGCTCTCGATGCCAAAGTCAAGCTGATTGCCAACATTTGTGCAACTCCCGC 1536
QY 1441 CAACTCTATGACCAATGATGATGACATGATGATCTGTGAGGAAATCTTCAGAACT 1500
Db 1537 CAACTCTATGACCAATGATGATGACATGATGATCTGTGAGGAAATCTTCAGAACT 1596
QY 1501 GGGCAAGACCTGCCAGGTCACCTCTGGAGCCCTGACCTGTGAGAGAGCGGCACC 1560
Db 1597 GGGCAAGACCTGCCAGGTCACCTCTGGAGCCCTGACCTGTGAGAGAGCGGCACC 1656
QY 1561 TACTACGTCTATGGATAGTGAAGTGGGCTGGAGTGTGGAGAGCCAGGGCTCTAC 1620
Db 1657 TACTACGTCTATGGATAGTGAAGTGGGCTGGAGTGTGGAGAGCCAGGGCTCTAC 1716
QY 1621 ACCCAAGTACCAATCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1680
Db 1717 ACCCAAGTACCAATCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1776
QY 1681 TAA 1683
Db 1777 TAA 1779
```

RESULT 4

AA45697
ID AAL45697 standard; DNA; 1683 BP.

XX AC AAL45697;

XX DT 13-JUN-2002 (first entry)

XX DE Human blood coagulation factor VII activating protease mutant DNA.

XX KW Human; blood coagulation factor VII activating protease; FSAP;

XX KW single-chain plasminogen activator; bleeding disorder; haematological;

XX KW haemostatic; mutant; gene; ds.

XX OS Homo sapiens.

XX OS Synthetic.

XX PH Key Location/Qualifiers

XX FT CDS 1..1683

XX FT /*cag= a

XX FT /product= "mutant FSAP"

XX FT

XX PN EP1182258-A1.

XX 27-FEB-2002.

XX 05-JUL-2001; 2001BP-00115691.

XX 26-JUL-2000; 2000DE-01036641.

XX 10-OCT-2000; 2000DE-01050040.

XX 21-OCT-2000; 2000DE-01052312.

XX 12-APR-2001; 2001DE-01018706.

XX (AVET) AVENTIS BEHRING GMBH.

XX Roemisch J, Stoeck H, Feussner A, Lang W, Weimer T, Becker M;

XX Nerlich C, Muth-Naumann G;

XX WPI; 2002-270939/32.

XX P-PSDB; AAO17145.

XX New nucleic acid encoding mutant factor 7 activating protease, useful for

XX diagnosis, treatment and prevention of coagulation disorders, also

XX related protein and antibodies.

XX Disclosure; Page 16-17; 27pp; German.

XX The present invention relates to a mutant of the DNA sequence encoding

XX the protease (FSAP) that activates blood coagulation factor VII (FVII)

XX and single-chain plasminogen activator, where at least one of the base

XX changes G to C at nucleotide 1177 and G to A at nucleotide 1601 is

XX present. The mutant sequences can be used in the treatment and prevention

XX of bleeding disorders associated with inherited or acquired defects of

XX blood coagulation factors V, VIII, IX, X, XI, XII, von Willebrand's

XX factor and/or with antibodies against any of these proteins. The present

XX sequence is the mutant human FSAP coding sequence

XX SQ Sequence 1683 BP; 441 A; 437 C; 435 G; 370 T; 0 U; 0 Other;

XX Query Match 99.8%; Score 1679.8; DB 6; Length 1683;

XX Best Local Similarity 99.8%; Pred. No. 0;

XX Matches 1681; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATGTTTCCAGGATGTCTGATCTCCATGTTCTCTCTTAATGCTCTGTGGGAAAGACA 60

Db 1 ATGTTTCCAGGATGTCTGATCTCCATGTTCTCTCTTAATGCTCTGTGGGAAAGACA 60

QY 61 GCCTGTGGTCTCCCTGATGCTTTTATGGAAGCTGGACCCAGCTGGACCCCTGAC 120

Db 61 GCCTGTGGTCTCCCTGATGCTTTTATGGAAGCTGGACCCAGCTGGACCCCTGAC 120

QY 121 CAGTATGATTACAGCTACGAGGATTTAATCAGGAAGAGAACACCACTAGTACACTTACC 180

Db 121 CAGTATGATTACAGCTACGAGGATTTAATCAGGAAGAGAACACCACTAGTACACTTACC 180

QY 181 CATGCTGAGAACTCTGATCTGATCTACACTGAGAGCAAGCTGATCCATCCAGCCCAAC 240

Db 181 CATGCTGAGAACTCTGATCTGATCTACACTGAGAGCAAGCTGATCCATCCAGCCCAAC 240

QY 241 CCCTGTGAACAGCTGGGAGTCTGCTCTCATGGGAGACCTTCACTGAGCTGCTG 300

Db 241 CCCTGTGAACAGCTGGGAGTCTGCTCTCATGGGAGACCTTCACTGAGCTGCTG 300

QY 301 GCTCTTTCTCTGGGAATAAGTGTGAGAAAGTCAAAATACGTGCAAGGACCACTATGT 360

Db 301 GCTCTTTCTCTGGGAATAAGTGTGAGAAAGTCAAAATACGTGCAAGGACCACTATGT 360

QY 361 GCGGGGGCAATGCTCTCATTTACCCAGAGTCTCCCTACTACGCTGTCTGTCTGTAACAC 420

Db 361 GCGGGGGCAATGCTCTCATTTACCCAGAGTCTCCCTACTACGCTGTCTGTCTGTAACAC 420

QY 421 CTTTACAGAGTCCAGCTGCTCCCAAGTGTCTCTTATGAGGCAAAACCCCTGCGAG 480

Db 421 CTTTACAGAGTCCAGCTGCTCCCAAGTGTCTCTTATGAGGCAAAACCCCTGCGAG 480

QY 481 AATGGGCTACTGCTCCCGCATAGGGAGATCCAAAGTTCACCTGTGCTGCTCCGAC 540

Db	1561	TACTACGTCATATCGGGATAGTGAAGCTGGGCGCTGGAGTGTGAGAAAGGCCAGGGGTCCTAC	1620
Qy	1621	ACCAAGATTACCAAAATTCCTGAATTGGATCAAAAGCCACCAATCAAAAGTGAAGTGGCTTC	1680
Db	1621	ACCAAGTTACCAAAATTCCTGAATTGGATCAAAAGCCACCAATCAAAAGTGAAGTGGCTTC	1680
Qy	1661	TAA 1683	
Db	1681	TAA 1683	
RESULT 5			
AAC76693			
ID	AAC76693	standard; cDNA; 3623 BP.	
XX	AAC		
XX	AAC		
XX	AAC		
XX	AAC		
DT	08-FEB-2001	(first entry)	
XX	Human ORFX	ORF2248 polynucleotide sequence SEQ ID NO:4495.	
DE	Human;	open reading frame; ORFX; detection; cytosstatic; hepatotropic;	
XX	Human;	open reading frame; ORFX; detection; cytosstatic; hepatotropic;	
KW	vulnary;	antipsoiatric; antiparkinsonian; noctropic; neuroprotective;	
KW	anticonvulsant;	osteopathic; antiarthritic; immunosuppressant; cardiac;	
KW	immunostimulant;	thrombolytic; coagulant; vasotropic; antidiabetic;	
KW	hypotensive;	dermatological; immunosuppressive; antiinflammatory;	
KW	antiviral;	antibacterial; antifungal; antirheumatic; antithyroid;	
KW	antianaemic;	gene therapy; cancer; proliferative disorder; hypertension;	
KW	neurodegenerative disorder;	osteoarthritis; graft vs host disease;	
KW	cardiovascular disease;	diabetes mellitus; hypothyroidism; SCID; AIDS;	
KW	cholesterol ester storage;	systemic lupus erythematosus; infection;	
KW	severe combined immunodeficiency;	malaria; autoimmune disorder; asthma;	
KW	allergy; aplastic anaemia;	nocturnal haemoglobinuria; burn; wound;	
KW	bone damage;	cartilage damage; antiinflammatory disease; coagulation;	
KW	thrombosis;	contraceptive; ss.	
XX	Homo sapiens.		
OS	WO200058473-A2.		
PN	05-OCT-2000.		
PD	31-MAR-2000;	2000WO-US008621.	
XX	31-MAR-1999;	99US-0127607P.	
PR	02-APR-1999;	99US-0127636P.	
PR	05-APR-1999;	99US-0127728P.	
PR	30-MAR-2000;	2000US-00540763.	
XX	(CURA-) CURAGEN CORP.		
PA	Shinkets RA,	Leach M;	
PI	WPI;	2000-602362/57.	
DR	P-PSDB;	AAB42484.	
DR	Novel nucleic acids and peptides derived from open reading frame X,		
PT	useful for treating e.g. cancers, proliferative disorders,		
PT	neurodegenerative disorders and cardiovascular disease.		
XX	Claim 5;	Page 3679-3681; 5507pp; English.	
PS	AAC74446 to AAC77606	encode the proteins given in AAB40237 to AAB43397,	
XX	which represent the human ORFX open reading frames 1 to 3161. The ORFX		
CC	sequences have activities such as: cytosstatic; hepatotropic; vulnary;		
CC	antipsoiatric; antiparkinsonian; noctropic; neuroprotective; osteopathic;		
CC	anticonvulsant; antiarthritic; immunosuppressant; immunostimulant;		
CC	cardiant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive;		
CC	dermatological; immunosuppressive; antirheumatic; antithyroid;		
CC	antiviral; antifungal; antirheumatic; antithyroid; and antianaemic. The		
CC	sequences can be used for determining the presence of or predisposition		
CC	to, or preventing or treating pathological conditions associated with an		
CC	ORFX-associated disorder. The nucleic acids can be used to express ORFX		

CC proteins in gene therapy vectors. The proteins and nucleic acids may be
CC used to treat cancers, proliferative disorders, neurodegenerative
CC disorders, osteoarthritis, graft vs host disease, cardiovascular disease,
CC diabetes mellitus, hypertension, hypothyroidism, cholesterol ester
CC storage, systemic lupus erythematosus, severe combined immunodeficiency
CC (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune
CC disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and
CC cartilage damage, nocturnal haemoglobinuria, anti-inflammatory disease; to
CC enhance coagulation; to inhibit thrombosis; and as a contraceptive
XX
SQ Sequence 3623 BP; 848 A; 1005 C; 930 G; 840 T; 0 U; 0 Other;

Query Match 99.7%; Score 1678.2; DB 3; Length 3623;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1680; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATGTTGGCAGATGTCGATCTCATGTTCTGCTGTTAATGCTCTGTTGGGAAAGACA 60
DB 97 ATGTTGGCAGATGTCGATCTCATGTTCTGCTGTTAATGCTCTGTTGGGAAAGACA 156
QY 61 GCCTGGGGTTCCTCCGTGATGCTTTATGGAAGCTGGACCCAGAGCTGACCCCTGAC 120
DB 157 GCCTGGGGTTCCTCCGTGATGCTTTATGGAAGCTGGACCCAGAGCTGACCCCTGAC 216
QY 121 CAGTATGATTACAGCTACGAGGATTATATCAGGAAGAGAACACAGTAGCACACTTACC 180
DB 217 CAGTATGATTACAGCTACGAGGATTATATCAGGAAGAGAACACAGTAGCACACTTACC 276
QY 181 CATGCTGAGAACCTGACTGTTACTACACTGAGGACCAAGCTGATCCATGCGGCCAAC 240
DB 277 CAGCTGAGAACCTGACTGTTACTACACTGAGGACCAAGCTGATCCATGCGGCCAAC 336
QY 241 CCCTGTGAACCGTGGGAGCTGCTCGTCCATGGGAGCACCTTCACATGAGCTGCGCTG 300
DB 337 CCCTGTGAACCGTGGGAGCTGCTCGTCCATGGGAGCACCTTCACATGAGCTGCGCTG 396
QY 301 GCTCCTTCTCTGGGAATAAGTGTGAGAAAGTGCAAAATAGCTGCAAGGACAAACCCATGT 360
DB 397 GCTCCTTCTCTGGGAATAAGTGTGAGAAAGTGCAAAATAGCTGCAAGGACAAACCCATGT 456
QY 361 GCCCGGGGCAATGCTCTATTACCCAGAGTCTCTCCCTACTACCGCTGTCTGTAAACAC 420
DB 457 GSCCGGGGCAATGCTCTATTACCCAGAGTCTCTCCCTACTACCGCTGTCTGTAAACAC 516
QY 421 CTTTACAGAGTCCAGCTGCTCCCAAGTGTCTCTGATGAGGCCCAAAACCCCTGCCAG 480
DB 517 CTTTACAGAGTCCAGCTGCTCCCAAGTGTCTCTGATGAGGCCCAAAACCCCTGCCAG 576
QY 481 AATGGGCTACCTGCTCCCGGATAGCGGAGATCCAAAGTTCACCTGTGCTGCCGAC 540
DB 577 AATGGGCTACCTGCTCCCGGATAGCGGAGATCCAAAGTTCACCTGTGCTGCCGAC 636
QY 541 CAGTTCAAGGGAAATTCGTGAAATAGGTTCTGATGACTGCTATGTTGGGATGGCTAC 600
DB 637 CAGTTCAAGGGAAATTCGTGAAATAGGTTCTGATGACTGCTATGTTGGGATGGCTAC 696
QY 601 TCTTACCGAGGAATGATGATAGGACAGTCAACAGCATGCGTGCCTTACTGNACTCC 660
DB 697 TCTTACCGAGGAATGATGATAGGACAGTCAACAGCATGCGTGCCTTACTGNACTCC 756
QY 661 CACCTCCTCTTCGAGGAAATACAAATGTTTATGGAGATGCTGAAACCCATGGGATT 720
DB 757 CACCTCCTCTTCGAGGAAATACAAATGTTTATGGAGATGCTGAAACCCATGGGATT 816
QY 721 GGGGAACACAATTCGTGAGAAACCCAGATGGGACGAAAGCCCTGTGCTTTATTTAAA 780
DB 817 GGGGAACACAATTCGTGAGAAACCCAGATGGGACGAAAGCCCTGTGCTTTATTTAAA 876
QY 781 GTTACCAATGACAAGGTGAAATGGGAATACTGTGATGCTCAGCTGCTCAGCCGAGAC 840
DB 877 GTTACCAATGACAAGGTGAAATGGGNACTACTGTGATGCTCAGCTGCTCAGCCGAGAC 936
QY 841 GTTGCTCTCCAGAGAAAGCCCCACTGAGCCATCAACCAAGCTTCCGGGGTTGACTCC 900

DB 937 GTTGGCTACCCAGAGAAAGCCCCACTGAGCCATCAACCAAGCTTCCGGGGTTGACTCC 996
QY 901 TGTGGAAGACTGAGATAGCAGAGAGAAAGATCAAGAGAATCTATGGAGCTTTTAAGAC 960
DB 997 TGTGGAAGACTGAGATAGCAGAGAGAAAGATCAAGAGAATCTATGGAGCTTTTAAGC 1056
QY 961 ACGCGGGGACAGCACCCATCGGAGGCGTCCCTCCAGTCTCGCTGCTGACCATCTCC 1020
DB 1057 ACGCGGGGACAGCACCCATCGGAGGCGTCCCTCCAGTCTCGCTGCTGACCATCTCC 1116
QY 1021 ATGCCCCAGGGCCACTTCTGTGTGGGGCGCTGATCCACCCCTGCTGGGTGCTCACTGCT 1080
DB 1117 ATGCCCCAGGGCCACTTCTGTGTGGGGCGCTGATCCACCCCTGCTGGGTGCTCACTGCT 1176
QY 1081 GCCCACTGACCCGACATAAACCACAGACATCTAAAGTGGTGTAGGGACCCAGGACCTG 1140
DB 1177 GCCCACTGACCCGACATAAACCACAGACATCTAAAGTGGTGTAGGGACCCAGGACCTG 1236
QY 1141 AAGAAAGAGAAATTTTCATGAGCAGAGCTTTAGGGTGGAGAAGATATTCAGGTACACCCAC 1200
DB 1237 AAGAAAGAGAAATTTTCATGAGCAGAGCTTTAGGGTGGAGAAGATATTCAGGTACACCCAC 1296
QY 1201 TACAATGAAGAGATGAGATTCCCAATGATATTTGCTCAAGTTAAAGCCAGTG 1260
DB 1297 TACAATGAAGAGATGAGATTCCCAATGATATTTGCTCAAGTTAAAGCCAGTG 1356
QY 1261 GATGGTCACTGTGCTCTAGAAATCCAAATACGTGAAGACTGTGTGCTTGTGCTGCTGCTCC 1320
DB 1357 GATGGTCACTGTGCTCTAGAAATCCAAATACGTGAAGACTGTGTGCTTGTGCTGCTGCTCC 1416
QY 1321 TTTCCCTCTGGGAGTGAAGTCCACATCTCTGGCTGGGGTGTTCACAGAAACAGGAAAGGG 1380
DB 1417 TTTCCCTCTGGGAGTGAAGTCCACATCTCTGGCTGGGGTGTTCACAGAAACAGGAAAGGG 1476
QY 1381 TCCCGCCAGCTCTCGATGCTCCAAAGTCAAGCTGATTCGCAACACTTTGTGCACTCCCGC 1440
DB 1477 TCCCGCCAGCTCTCGATGCTCCAAAGTCAAGCTGATTCGCAACACTTTGTGCACTCCCGC 1536
QY 1441 CAACTCTATGACACACATGATTTGATGACAGTATGATCTGTGCAAGAAATCTTCAGAAACCT 1500
DB 1537 CAACTCTATGACACACATGATTTGATGACAGTATGATCTGTGCAAGAAATCTTCAGAAACCT 1596
QY 1501 GGGCAGACACCTGCCAGGGTGAATCTGGAGCCCCCTGACCTGTGAGAGAGCGGACCC 1560
DB 1597 GGGCAGACACCTGCCAGGGTGAATCTGGAGCCCCCTGACCTGTGAGAGAGCGGACCC 1656
QY 1561 TACTAGCTCTATGGGATAGTGAAGTGGGGCTGGAGTGTGGGAAGAGGCCAGGGGTCTAC 1620
DB 1657 TACTAGCTCTATGGGATAGTGAAGTGGGGCTGGAGTGTGGGAAGAGGCCAGGGGTCTAC 1716
QY 1621 ACCCAAGTTACCAAAATTCCTGAATTCGAATCAAAAGCCACCATCAAAAGTGAAGTGGCTTC 1680
DB 1717 ACCCAAGTTACCAAAATTCCTGAATTCGAATCAAAAGCCACCATCAAAAGTGAAGTGGCTTC 1776
QY 1681 TAA 1683
DB 1777 TAA 1779

RESULT 6
AAA44763
ID AAA44763 standard; cDNA; 617 BP.

XX AAA44763;

XX 21-AUG-2000 (first entry)

XX Human secreted expressed sequence tag SEQ ID NO:1338.

XX Human; mouse; chicken; rat; secreted expressed sequence tag; sEST;
XX expressed sequence tag; sST; probe; chemotactic; proliferative;
XX immunomodulatory; haematopoietic; chemokinetic; analgesic; haemostatic;

thrombolytic; antiinflammatory; cytostatic; antibacterial; antifungal; antiviral; antidiabetic; antiasthmatic; vulnary; antiparkinsonian; anticancer; osteoprotective; neuroprotective; neurotropic; antipsoriatic; cerebroprotective; anticonvulsant; antidepressant; gene therapy; vaccine; autoimmune disorder; multiple sclerosis; allergic condition; insulin dependent diabetes; asthma; myeloid cell deficiency; ulcer; lymphoid cell deficiency; burn; osteoporosis; osteoarthritis; central nervous system disorder; Alzheimer's disease; stroke; Parkinson's disease; Huntington's disease; coagulation disorder; haemophilia; thrombosis; inflammatory disorder; Crohn's disease; tumour; infection; depression; psoriasis; ss.

Homo sapiens.
 WO200021991-A1.
 20-APR-2000.
 15-OCT-1999; 99WO-US024206.
 15-OCT-1998; 98US-0104436P.
 (GEMY) GENETICS INST INC.
 Jacobs K, McCoy JM, Lavallie ER, Collins-Racie LA, Evans C;
 Merberg D, Treacy M, Bowman MR;
 WPI; 2000-317938/27.

Isolated polynucleotides, and encoded proteins, comprising secreted expressed sequence tags (SESTs), useful for treating various disorders such as autoimmune, infectious, and central nervous system disorders.

Claim 1; Page 549; 803pp; English.
 AAA43426 to AAA45925 represent specifically claimed secreted expressed sequence tags (SESTs), isolated from human, mouse, chicken and rat tissue sources. The SESTs can have a range of activities depending on the tissues they were isolated from. The activities include: chemotactic; proliferative; immunomodulatory; haematopoietic; chemokinetic; analgesic; haemostatic; thrombolytic; antiinflammatory; cytostatic; antibacterial; antifungal; antiviral; antidiabetic; antiasthmatic; vulnary; anticancer; osteoprotective; neuroprotective; neurotropic; antiparkinsonian; antipsoriatic; cerebroprotective; anticonvulsant; and antidepressant. The SESTs can be used for gene therapy and in vaccines. The SESTs are useful as probes for the identification and isolation of full-length cDNAs and genomic DNA molecules which correspond to the SESTs. Proteins encoded by the SESTs are useful in assays for determining biological activity and raising antibodies. They may be useful for treatment of autoimmune disorders (multiple sclerosis, insulin dependent diabetes), allergic conditions (asthma), myeloid or lymphoid cell deficiencies, wounds, burns, ulcers, osteoporosis, osteoarthritis, central nervous system disorders (Alzheimer's, Parkinson's, Huntington's disease, stroke), coagulation disorders (haemophilia, thrombosis), inflammatory disorders (Crohn's disease), tumours, bacterial, fungal or viral infections, depression and psoriasis. AAA45926 to AAA45931 represent linker variants which are given in the exemplification of the present invention

Sequence 617 BP; 165 A; 169 C; 148 G; 135 T; 0 U; 0 Other;
 Query Match 31.3%; Score 526.4; DB 3; Length 617;
 Best Local Similarity 99.8%; Pred. No. 3.1e-140;
 Matches 527; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

1 ATGTTTGCAGAGTCTGATCTCCATGTTCTGCTGTTATGCTGTTGGGAAAGACA 60
 84 ATGTTTGCAGAGTCTGATCTCCATGTTCTGCTGTTATGCTGTTGGGAAAGACA 143
 61 GCCTGTGGGTTCTCCTGATGCTTTTATGGAAGCTGGACCCAGACTGGACCCCTGAC 120
 144 GCCTGTGGGTTCTCCTGATGCTTTTATGGAAGCTGGACCCAGACTGGACCCCTGAC 203
 121 CAGTATGATTCAGCTACGAGGATTAATCAGGAGAGAACACCACTAGCACACTTACC 180

204 CAGTATGATTCAGCTACGAGGATTAATCAGGAGAGAACACCACTAGCACACTTACC 263
 181 CATCTGAGATCTGATCTGTTACTACACTGAGACCAAGCTGATCCATCCAGCCCAAC 240
 264 CACGCTGAGATCTGATCTGTTACTACACTGAGACCAAGCTGATCCATCCAGCCCAAC 323
 241 CCCTGTGAACACGGTGGGACTGCTGCTCATGGGAGACCTTCACATCAGCTGCTG 300
 324 CCCTGTGAACACGGTGGGACTGCTGCTCATGGGAGACCTTCACATCAGCTGCTG 383
 301 GCTCTTTCTCTGGGAATAAGTGTGAGAAAGTGCAGAAATACGTGCAAGGACAAACCCATGT 360
 384 GCTCTTTCTCTGGGAATAAGTGTGAGAAAGTGCAGAAATACGTGCAAGGACAAACCCATGT 443
 361 GCGCGGGCCAAATGCTCTCATCCAGAGTCTCCTACTACTACCGCTGCTGTGTAACAC 420
 444 GCGCGGGCCAAATGCTCTCATCCAGAGTCTCCTACTACTACCGCTGCTGTGTAACAC 503
 421 CCTTACACAGTCCAGCTGCTCCCAAGTGTCTCTGTATGCAAGCCAAACCCCTGCGAG 480
 504 CCTTACACAGTCCAGCTGCTCCCAAGTGTCTCTGTATGCAAGCCAAACCCCTGCGAG 563
 481 AATGGGGTCTACTGCTCCCGCATAGCGGAGATCCCAAGTTCACCTGT 528
 564 AATGGGGTCTACTGCTCCCGCATAGCGGAGATCCCAAGTTCACCTGT 611

RESULT 7

ACH21030

ID ACH21030 standard; cDNA; 428 BP.

XX ACH21030;

XX 13-OCT-2003 (first entry)

XX Human adult liver cDNA #642.

DE Human; ss; sequencing by hybridisation; SBH; expressed sequence tag; EST;
 KW genome mapping; biodiversity; genetic disorder.

OS Homo sapiens.

XX US2003073623-A1.

PD 17-APR-2003.

XX 30-JUL-2001; 2001US-00918995.

XX 30-JUL-2001; 2001US-00918995.

XX (DRNA/) DRMANAC R T.

XX (LABA/) LABAT I.

XX (STAC/) STACHE-CRAIN B.

XX (DICK/) DICKSON M C.

XX (JONE/) JONES L W.

XX Drmanac RT, Labat I, Stache-Crain B, Dickson MC, Jones LW;

XX WPI; 2003-615964/58.

XX New polynucleotide sequences obtained from various cDNA libraries, useful
 PT as hybridization probes, as oligomers for PCR, for chromosome and gene
 PT mapping, in the recombinant production of protein, or in generating
 PT antisense DNA or RNA.

XX Claim 1; SEQ ID NO 8242; 44pp; English.

XX The invention relates to an isolated polynucleotide comprising any one of
 CC 38043 cDNA sequences, appearing as ACH12789-ACH50831, whose sequence was
 CC determined by the technique of SBH (sequencing by hybridisation). Also
 CC included is a purified polypeptide comprising a sequence corresponding to
 CC a reading frame of the novel polynucleotide. The nucleic acid sequences

are useful in diagnostics as expressed sequence tags (EST) for
 CC identifying expressed genes or for physical mapping of the human genome,
 CC in forensics, in assessing biodiversity, or in identifying mutations
 CC responsible for genetic disorders and other traits. The nucleotide
 CC sequences are also useful as hybridisation probes, as oligomers for PCR,
 CC for chromosome and gene mapping, in the recombinant production of
 CC protein, or in generating antisense DNA or RNA. The purified polypeptide
 CC is useful for generating antibodies specific for it. The present sequence
 CC is one of the 38043 isolated cDNA/EST sequences. Note: The sequence data
 CC for this patent did not form part of the printed specification, but was
 CC obtained in electronic format directly from USPTO at
 CC seqdata.uspto.gov/sequence.html?docID=20030073623
 SQ
 Query Match 21.9%; Score 368.4; DB 8; Length 428;
 Best Local Similarity 99.7%; Pred. No. 5.2e-95;
 Matches 369; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 ATGTTTGCCAGGATGCTGATCTCCATGTTCTGCTGTTAATGGCTCTGGTGGGAAAGACA 60
 Db 59 ATGTTTGCCAGGATGCTGATCTCCATGTTCTGCTGTTAATGGCTCTGGTGGGAAAGACA 118
 QY 61 GCTCTGGGTTCTCCCTGATGCTTTATTGGAAAGCTTGACCCAGACTGGACCCCTGAC 120
 Db 119 GCTCTGGGTTCTCCCTGATGCTTTATTGGAAAGCTTGACCCAGACTGGACCCCTGAC 178
 QY 121 CAGTATGATTACAGCTACGAGGATTATATACAGGAGAGAACACACAGTAGCACACTTACC 180
 Db 179 CAGTATGATTACAGCTACGAGGATTATATACAGGAGAGAACACACAGTAGCACACTTACC 238
 QY 181 CATGCTGAGAACTCTGACTGTTACTACTGAGGACCAAGCTGATCCATGCCAGCCCAAC 240
 Db 239 CATGCTGAGAACTCTGACTGTTACTACTGAGGACCAAGCTGATCCATGCCAGCCCAAC 298
 QY 241 CCTGTGAACACGGTGGGACTGCTCTGCCAGGACACCTTACATGACAGCTGCCTG 300
 Db 299 CCTGTGAACACGGTGGGACTGCTCTGCCAGGACACCTTACATGACAGCTGCCTG 358
 QY 301 GCTCCTTTCTCTGGGAATAGTGTGAGAAAGTGCAGAAATACGTGCAAGGACAAACCATGT 360
 Db 359 GCTCCTTTCTCTGGGAATAGTGTGAGAAAGTGCAGAAATACGTGCAAGGACAAACCATGT 418
 QY 361 GGCCGGGGCC 370
 Db 419 GGCCGGGGCC 428
 RESULT 8
 ID ABI9281
 XX ABI9281 standard; cDNA; 397 BP.
 AC ABI9281;
 XX
 XX 07-MAR-2002 (first entry)
 XX
 XX Mouse ischaemic condition related cDNA sequence SEQ ID NO:123.
 DE
 XX Mouse; ischaemia; compressive ischaemia; occlusive ischaemia;
 KW vasospastic ischaemia; ischaemic condition; ischaemic disease; ss.
 XX
 XX Mus musculus.
 OS
 XX WO200188188-A2.
 EN
 XX 22-NOV-2001.
 PD
 XX 18-MAY-2001; 2001WO-JP004192.
 PF
 XX 18-MAY-2000; 2000JP-00145977.
 PR
 XX (UTNI-) UNIV NIHOON SCHOOL JURIDICAL PERSON.
 PA
 XX

Ishikawa K, Asai S, Takahashi Y, Nagata T, Ishii Y;
 WPI; 2002-034733/04.
 XX Examining the ischemic condition (e.g. occlusive ischemia) by measuring
 PT expression levels of particular genes defined in the specification or by
 PT determining the expression profile of a gene group comprising these
 PT genes.
 XX
 PS Claim 2; Page 351-352; 2690pp; English.
 XX
 CC The present invention describes a method for examining ischaemic
 CC conditions, comprising measuring the expression levels of particular
 CC genes (I) in a test sample or determining the expression profile of a
 CC gene group in the sample comprising genes selected from (I). The method
 CC is useful for examining the ischaemic condition (e.g. compressive
 CC ischaemia, occlusive ischaemia or vasospastic ischaemia) by measuring
 CC expression levels of particular genes (ABI9281 to ABI9912, encoding the
 CC protein sequences in ABB57020 to ABB57374) or by determining the
 CC expression profile of a gene group comprising these genes. The expression
 CC levels or expression profiles produced by these genes are used as an
 CC indicator when screening for ischaemic condition-improving drugs or
 CC therapeutics for ischaemic diseases. ABI9913 and ABI9914 represent PCR
 CC primers for a mouse ischaemic condition related sequence, which are used
 CC in the exemplification of the present invention
 XX
 SQ Sequence 397 BP; 103 A; 105 C; 111 G; 78 T; 0 U; 0 Other;
 Query Match 14.4%; Score 242.6; DB 6; Length 397;
 Best Local Similarity 84.7%; Pred. No. 6e-59;
 Matches 272; Conservative 0; Mismatches 49; Indels 0; Gaps 0;
 QY 1363 ACAGAAACAGAAAGGTCCTCCAGCTCTCTGGATGCCAAAGTCAAGCTGATGTCGAAC 1422
 Db 13 ACAGAAACAGAAAGGTCCTCCAGCTCTCTGGATGCCAAAGTCAAGCTGATGTCGAAC 72
 QY 1423 ACTTTGTGCAACTCCGCCAACTCTATGACCAATGATGATGACAGTATGATCTGTGCA 1482
 Db 73 CTTTGTGCAACTCCGCCAACTCTATGACCAATGATGATGATGATGATGTCGG 132
 QY 1483 GGAAATCTTCAGAAACCTGGGAGGACCTGCCAGGTCACCTCTGGAGCCCTGAC 1542
 Db 133 GGGAACTCTCAGAAAGCCGGATCAGACCTTCCAGGGTACCTCGGGGGCCCTCTAAC 192
 QY 1543 TGTGAGAAGGACGCGACCTACTACTGCTATGGATAGTACGCTGGGCTGGAGTGTGG 1602
 Db 193 TGTGAGAAGGATGGAACCTTACTACTGCTACGGGATGTAAGCTGGGCGGAGTGTGG 252
 QY 1603 AAGAGCCAGGGTCTACACCCAAAGTTACCAATTCCTGAATTCGATCAAGCCCATC 1662
 Db 253 AAGAGCCAGGAGTCTACACTCAAGTCAAGTCAAGTTCCTGAATTCGATCAAGCCCATC 312
 QY 1663 AAAAGTGAAAGTGGCTTCTAA 1683
 Db 313 CACAGGGAGGCTGGCCTCTGA 333
 RESULT 9
 ID ABA58323/c
 XX ABA58323 standard; DNA; 451 BP.
 XX
 AC ABA58323;
 XX
 XX 01-FEB-2002 (first entry)
 DT
 XX Human foetal liver single exon nucleic acid probe #6628.
 DE
 XX Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
 KW
 XX Homo sapiens.
 OS
 XX WO200157277-A2.
 EN
 XX

PR 26-MAY-2000; 2000US-0207456P.
 PR 30-JUN-2000; 2000US-00608408.
 PR 03-AUG-2000; 2000US-00632366.
 PR 21-SEP-2000; 2000US-0234687F.
 PR 27-SEP-2000; 2000US-0236359P.
 PR 04-OCT-2000; 2000GB-00024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX
 XX WPI; 2001-488897/53.
 XX
 XX Human genome-derived single exon nucleic acid probes useful for analyzing
 PT gene expression in human placenta.
 XX
 PS Claim 25; SEQ ID NO 6633; 654pp; English.
 XX
 CC The present invention relates to single exon nucleic acid probes (SENP).
 CC The present sequence is one such probe. The probes are useful for
 CC producing a microarray for predicting, measuring and displaying gene
 CC expression in samples derived from human placenta. The probes are useful
 CC for antenatal diagnosis of human genetic disorders
 XX
 XX Sequence 451 BP; 116 A; 102 C; 127 G; 106 T; 0 U; 0 Other;
 SQ
 Query Match 7.1%; Score 120.2; DB 4; Length 451;
 Best Local Similarity 97.6%; Pred. No. 8.1e-24;
 Matches 122; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 324 TCAGAAGTGC AAAATACGTGCAAGGACAAACCCATGTGGCGGGGCCAATGTCTCATTTAC 383
 DB 451 TCCTACAGTGCA AAAATACGTGCAAGGACAAACCCATGTGGCGGGGCCAATGTCTCATTTAC 392
 QY 384 CCAGAGTCTCTCCCTACTACCGTGTGTCTGTATAACACCTTTACAGGTCCTCAGTGTCTC 443
 DB 391 CCAGAGTCTCTCCCTACTACCGTGTGTCTGTATAACACCTTTACAGGTCCTCAGTGTCTC 332
 QY 444 CCAAG 448
 DB 331 CCAAG 327
 RESULT 11
 AAK32096/C
 ID AAK32096 standard; DNA; 451 BP.
 XX
 AC AAK32096;
 DT 06-NOV-2001 (first entry)
 XX Human bone marrow expressed single exon probe SEQ ID NO: 6653.
 DE
 XX Human; bone marrow expressed exon; gene expression analysis; probe;
 KW microarray; cancer; leukaemia; lymphoma; myeloma; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200157276-A2.
 XX
 PD 09-AUG-2001.
 XX
 XX 30-JAN-2001; 2001WO-US000868.
 XX
 PR 04-FEB-2000; 2000US-0180312P.
 PR 26-MAY-2000; 2000US-0207456P.
 PR 30-JUN-2000; 2000US-00608408.
 PR 03-AUG-2000; 2000US-00632366.
 PR 21-SEP-2000; 2000US-0234687P.
 PR 27-SEP-2000; 2000US-0236359P.
 PR 04-OCT-2000; 2000GB-00024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488900/53.
XX Human genome-derived single exon nucleic acid probes useful for analyzing
PT gene expression in human bone marrow.
XX
XX Example 4; SEQ ID NO 6653; 658pp + Sequence Listing; English.
XX
XX The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC bone marrow. They can be used to measure gene expression in bone marrow
CC samples, which may enable the improved diagnosis and treatment of cancers
CC such as lymphoma, leukaemia and myeloma. The present sequence is one of
CC the probes of the invention
XX
XX Sequence 451 BP; 116 A; 102 C; 127 G; 106 T; 0 U; 0 Other;
SQ
Query Match 7.1%; Score 120.2; DB 4; Length 451;
Best Local Similarity 97.6%; Pred. No. 8.1e-24;
Matches 122; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 324 TCAGAAAGTGCAAAATACGTGCAAGGACAAACCCATGTGGCCGGGCCCAATGTCATTAC 383
Dd 451 TCTACAGTGCAAAATACGTGCAAGGACAAACCCATGTGGCCGGGCCCAATGTCATTAC 392
QY 384 CCAGAGTCTCCCTACTACCGCTGTGTGTGTAAACACCCCTTACACAGGTCCTGCTC 443
Dd 391 CCAGAGTCTCCCTACTACCGCTGTGTGTGTAAACACCCCTTACACAGGTCCTGCTC 332
QY 444 CCAAG 448
Dd 331 CCAAG 327

RESULT 12
AAK06420/c
ID AAK06420 standard; DNA; 451 BP.
XX
AC AAK06420;
XX
XX 05-NOV-2001 (first entry)
DT
XX Human brain expressed single exon probe SEQ ID NO: 6411.
DE
XX Human; brain expressed exon; gene expression analysis; probe; microarray;
KW Alzheimer's disease; multiple sclerosis; schizophrenia; epilepsy; cancer;
KW ss.
XX
XX Homo sapiens.
OS
XX WO200157275-A2.
PN
XX 09-AUG-2001.
PD
XX 30-JAN-2001; 2001WO-US0000667.
PF
XX 04-FEB-2000; 2000US-0180312P.
PR
XX 26-MAY-2000; 2000US-0207456P.
PR
XX 30-JUN-2000; 2000US-00608408.
PR
XX 03-AUG-2000; 2000US-00632366.
PR
XX 21-SEP-2000; 2000US-0234687P.
PR
XX 27-SEP-2000; 2000US-0236359P.
PR
XX 04-OCT-2000; 2000GB-00024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
PA
XX Penn SG, Hanzel DK, Chen W, Rank DR;
PI
XX WPI; 2001-483446/52.
XX
XX Single exon nucleic acid probes for analyzing gene expression in human

PT brains.
XX
XX Example 4; SEQ ID NO 6411; 650pp + Sequence Listing; English.
XX
XX The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC brain. They can be used to measure gene expression in brain cell samples,
CC which may enable the diagnosis and improved treatment of nervous system
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
CC epilepsy and cancers. The present sequence is one of the probes of the
CC invention
XX
XX Sequence 451 BP; 116 A; 102 C; 127 G; 106 T; 0 U; 0 Other;
SQ
Query Match 7.1%; Score 120.2; DB 4; Length 451;
Best Local Similarity 97.6%; Pred. No. 8.1e-24;
Matches 122; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 324 TCAGAAAGTGCAAAATACGTGCAAGGACAAACCCATGTGGCCGGGCCCAATGTCATTAC 383
Dd 451 TCTACAGTGCAAAATACGTGCAAGGACAAACCCATGTGGCCGGGCCCAATGTCATTAC 392
QY 384 CCAGAGTCTCCCTACTACCGCTGTGTGTGTAAACACCCCTTACACAGGTCCTGCTC 443
Dd 391 CCAGAGTCTCCCTACTACCGCTGTGTGTGTAAACACCCCTTACACAGGTCCTGCTC 332
QY 444 CCAAG 448
Dd 331 CCAAG 327

RESULT 13
ABS31791/c
ID ABS31791 standard; DNA; 451 BP.
XX
AC ABS31791;
XX
XX 25-FEB-2003 (first entry)
DT
XX Human liver single exon probe, SEQ ID No 6781.
DE
XX Human; single exon nucleic acid probe; liver; cirrhosis;
KW hyperlipoproteinaemia; hyperlipidaemia; hypercholesterolaemia;
KW coronary heart disease; ss.
XX
XX Homo sapiens.
OS
XX WO200157273-A2.
PN
XX 09-AUG-2001.
PD
XX 30-JAN-2001; 2001WO-US0000664.
PF
XX 04-FEB-2000; 2000US-0180312P.
PR
XX 26-MAY-2000; 2000US-0207456P.
PR
XX 30-JUN-2000; 2000US-00608408.
PR
XX 03-AUG-2000; 2000US-00632366.
PR
XX 21-SEP-2000; 2000US-0234687P.
PR
XX 27-SEP-2000; 2000US-0236359P.
PR
XX 04-OCT-2000; 2000GB-00024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
PA
XX Penn SG, Hanzel DK, Chen W, Rank DR;
PI
XX WPI; 2001-488898/53.
XX
XX Human genome-derived single exon nucleic acid probes useful for analyzing
PT gene expression in human adult liver.
XX
XX Claim 1; SEQ ID NO 6781; 658pp; English.
PS
XX The invention relates to a single exon nucleic acid probe (SENP) (I) for

CC	measuring human gene expression in a sample derived from human adult
CC	liver, comprising one of 13109 defined nucleotide sequences given in the
CC	specification (or complements/ fragments). The probe hybridises at high
CC	stringency to a nucleic acid molecule expressed in the human adult liver.
CC	(I) may be used for predicting, measuring and displaying gene expression
CC	in samples derived from human adult liver. The genes identified may be
CC	involved in genetic liver diseases such as cirrhosis,
CC	hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which is
CC	associated with coronary heart disease. ABS25011-ABS1005 represent human
CC	liver single exon nucleic acid probes of the invention. Note: The
CC	sequence information for this patent does not appear in the printed
CC	specification but was obtained in electronic format directly from WIPO at
CC	ftp.wipo.int/pub/published_pct_sequences
XX	
SQ	Sequence 451 BP; 116 A; 102 C; 127 G; 106 T; 0 U; 0 Other;
	Query Match 7.1%; Score 120.2; DB 4; Length 451;
	Best Local Similarity 97.6%; Pred. No. 8.le-24;
	Matches 122; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY	324 TCAGAACTGC AAAATACGTGTCAGGACACACCATTGGCGCGGGGCCAATGTCCTATTAC 383
Ddb	451 TCCTACAGTGC AAAATACGTGTCAGGACACACCATTGGCGCGGGGCCAATGTCCTATTAC 392
QY	384 CCAGAGTGCTCCTACTACCGCTGTGCTGTATAACACCCCTTACACAGGTCCCAGGTGTC 443
Ddb	391 CCAGAGTGCTCCTACTACTACCGCTGTGCTGTATAACACCCCTTACACAGGTCCCAGGTGTC 332
QY	444 CCAAG 448
Ddb	331 CCAAG 327

RESULT 14	
ABS06963/C	
ID	ABS06863 standard; DNA; 451 BP.
XX	
XX	
XX	ABS06863;
XX	AC
XX	
DT	19-AUG-2002 (first entry)
XX	
DE	Human genome-derived single exon probe from lung SEQ ID NO 6954.
XX	
XX	Human; ds; single exon probe; asthma; lung cancer; COPD; ILD;
KW	chronic obstructive pulmonary disease; interstitial lung disease;
KW	familial idiopathic pulmonary fibrosis; neurofibromatosis;
KW	tuberosus sclerosis; Gaucher's disease; Niemann-Pick disease;
KW	Hernandez-Fadlak syndrome; sarcoidosis; pulmonary haemosiderosis;
KW	pulmonary histiocytosis; lymphangioleiomyomatosis; Karagenar syndrome;
KW	pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;
KW	primary ciliary dyskinesia; pulmonary hypertension;
KW	hyaline membrane disease.

OS	Homo sapiens.
XX	
PN	WC200186003-A2.
XX	
XX	
PD	15-NOV-2001.
XX	
PF	
XX	30-JAN-2001; 2001WO-US000665.
XX	
PR	04-FEB-2000; 2000US-0180312P.
PR	26-MAY-2000; 2000US-0207456P.
PR	30-JUN-2000; 2000US-0060840B.
PR	03-AUG-2000; 2000US-00632366.
PR	21-SEP-2000; 2000US-023468P.
PR	27-SEP-2000; 2000US-0236359P.
XX	04-OCT-2000; 2000GB-00024263.
XX	
PA	(MOLE-) MOLECULAR DYNAMICS INC.
XX	
PI	Penn SG, Hanzel DK, Chen W, Rank DR;
XX	

XX Hepatocyte growth factor converting protease coding sequence.
DE
XX Hepatocyte growth factor; protease; cleavage; active, inactive;
KW precursor; ds.
XX
XX Homo sapiens.
OS
XX
XX Location/Qualifiers
FH 1.1968
FT /*tag= a
FT /product= "Hepatocyte growth factor converting protease."
XX
XX EP596524-A2.
PN
XX
XX 11-MAY-1994.
PD
XX
XX 05-NOV-1993; 93EP-00117988.
PF
XX
XX 05-NOV-1992; 92JP-00286133.
PR
XX 20-NOV-1992; 92JP-00312234.
PR
XX 20-NOV-1992; 92JP-00312242.
PR
XX (MITU) MITSUBISHI KASEI CORP.
XX
XX Shimomura T, Yamada K, Morimoto Y, Kitamura N, Miyazawa K;
PI
XX WPI; 1994-152921/19.
XX
XX P-PSDB; AAR53962.
DR
XX
XX Hepatocyte growth factor converting protease and precursor and gene
PT encoding them - for producing active two chain HGF from inactive single
PT chain HGF.
XX
XX Claim 15; Page 26-27; 30pp; English.
XX
XX The polypeptide encoded by this sequence has protease activity and is
CC capable of converting inactive single chain hepatocyte growth factor
CC (HGF) into active two chain HGF by cleavage at a specific site. (Updated
CC on 25-MAR-2003 to correct PN field.)
XX
XX
SQ Sequence 2033 BP; 339 A; 739 C; 623 G; 332 T; 0 U; 0 Other;
Query Match 7.0%; Score 117.6; DB 2; Length 2033;
Best Local Similarity 49.8%; Pred. No. 9.2e-23;
Matches 607; Conservative 0; Mismatches 559; Indels 54; Gaps 10;
QY 469 AACCCCTGCCAGATGGGGTACTCTGCTCCGCGCATAGGGGATCCAGTTCACCTGT 528
Db 742 AGCCCTTGCTGAACGGGGACCTGCCACCTGATCTGGCCACCGGACCCCGTGT 801
QY 529 GCCTGTCCCGACAGTTCAAGGGGAAATCTGTGAATAGTTCATGATGA---CTGCTAT 585
Db 802 GCCTGCCACACAGCTTCGCTGACGGCTCTGCAACATCGAGCTGATGAGCGCTGCTTC 861
QY 586 GTTGGCGATGGCTACTCTTACCGAGGGGAAATGAATAGGACATCAACGAGCATGGTGC 645
Db 862 TTGGGGAACGGCACTGGGTACCGTGGGTGGCGGACGACCTCAGCCTCGGGCCCTCAGTGC 921
QY 646 CTTTACTGGACCTCCACCTCTTTCAGGAGAAATTTACAAATGTTTATGGAGGATGCT 705
Db 922 CTGGCCTGGAACTCCGATCTGCTTACAGAGCTGCACTGGACTCCGTGGCGCGCGG 981
QY 706 GAAACCCATGGGATGGGGAACACAATTTCTGCAAGAACCCAGATGGGACGAAAGCCCC 765
Db 982 GCCCTGTGGCCCTGGGCCCCCATGCTTACTGCCGGAATCCGGACAATGACGAGAGGCC 1041
QY 766 TGGTGCTTTTAAAGTTACCAATGACAAGGTGAATGGGAATCTGTGATGCTCAGCC 825
Db 1042 TGGTGCTACGT---GGTGAAGGACAGCGCGCTCTCTCTGGGAGTACTGCCCGCTGGAGGCC 1098
QY 826 TGCTCAGCCACGAGCGTGGCTTACCCAGAGGAAAGCCCCCATAGAGCCATCAACCAAGCTT 885

Db 1099 TCGAATCCCT-----CACCAGAGTCCAAGTGTCAACCGATCTCTCGCGACCCCTG 1149
QY 886 CCGGGTTTGAATCTCTGTGAAAGACTGAGTAGCAGAGAGAGATCAAGAGATCTAT 945
Db 1150 CTTGAGCCAGCCTCCCGGGCGCCAGCCTCTCGCGAGGAGGACAAAGAGAGGACGTT 1209
QY 946 GGAGGCTTTAAGACGACGCGGGCAAGCAACCATGGCAGGGGTCTCCAGTCTCGTG 1005
Db 1210 CTGGCGCCAGGTATCATCGGCGCTCTCTCTCGTCCCGGCTCGCACCCCTG----- 1262
QY 1006 CTTCTGACCATCTCCATGCCCGGCGCACTTCTGTGTGGGGCGCTGATCCACCCCTGC 1065
Db 1263 --GCTGGCGCATCTACATCGGGGACAGCTTCTCGCGGGAGGCTGTGTCCACACCTGC 1320
QY 1066 TGGTGCTCACTGCTGCCCGACCTGACCGGACATAAAGAACAGACATCTAA-----AGGTG 1119
Db 1321 TGGGTGGTGTGGCGCCGCCCTCTCTTCAACCGACGAGCGAGCTGACGACCTTCGCGATCGAG 1380
QY 1120 GTCTAGGGGACCGAGGACCTGAAGAAAGAAATTTATGAGCAGAGCTTTAGGTGGAG 1179
Db 1381 GTCTGGCGCAGCACTTCTTCAACCGACGAGCGAGCTGACGACCTTCGCGATCGAG 1440
QY 1180 AAGATATTCAAGTACAGCCACTACATGAAGAGATGAGATTCGCCACAAATGATATGCA 1239
Db 1441 AAGTACATCCCGTACACCTGTACTCGGTGTTCAACCCGAGCGAC---GACCTCGTC 1497
QY 1240 TTGCTCAAGTTAAAGCCAGTGGATGTCTACTGTCTTGAATPCCAAATAGTGAAGACT 1299
Db 1498 CTGATCGGCTGAAGAAAGAGGACCGCTGTGCCACACGCTCGCAGTTCTGTGACGCC 1557
QY 1300 GTGTGCTGCTGAT-----GGGTCTTCTCCCTCTGGGAGTGGTCCACATCTCTGGC 1353
Db 1558 ATCTGCTGCCGAGCCCGGACGACCTTCCCGCAGGACACAGTGGCCAGATTGGGGC 1617
QY 1354 TGGGG-----TGTTACAGAAACAGGAAAGGGTCCCGCAGCTCTCTGGATGCCAAAGTC 1407
Db 1618 TGGGGCCACTTGGATGAGACGTGAGCGGCTACTCCAGCTCTCCGCGGAGGCGCTGGTC 1677
QY 1408 AAGTGTATGCCAACACTTTGTGCAACTCCCGCAACTCTATGACACACATGATGATGAC 1467
Db 1678 CCCCTGGTCCGCGACCAACAAGTGCAGCAGCCCTGAGGTCTACGGCGCGGACATCAGCCCC 1737
QY 1468 AGTATGATCTGTGAGGAAATCTTCAGAAACCTGGGCAAGACACCTGCCAGGGTGAATCT 1527
Db 1738 AACATGCTGTGCGGCTACTT---CGACTGCAAGTCCGACGCTGCCAGGGGACTCA 1794
QY 1528 GGAGGCGCCCTGACCTGTGAGAGGACGCGACCTTACTGCTCTATGGGATAGTGGTGG 1587
Db 1795 GGGGGGCGCCCTGGCTTCGAGAGAGAGCGGCTGGCTTACCTTACCGGATCATCAGCTGG 1854
QY 1588 GGCCTGGAGTGGG-----AAGAGCGCAGGGGTCTACACCCCAAGTTACCAAAATTCCTG 1641
Db 1855 GGTGACGGCTCGGGCGGCTCCACAGCGGGGTCTACACCCGCGTGGCACTATGTG 1914
QY 1642 AATTGGATCAAGACCCAT 1661
Db 1915 GACTGGATCAAGACCGAT 1934

Search completed: May 25, 2004, 06:30:47
Job time : 479 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 25, 2004, 05:42:15 ; Search time 3059 Seconds
(without alignments)
16429.568 Million cell updates/sec

Title: US-09-912-559-1
Perfect score: 1683
Sequence: 1 atgtgttcaggatgtctga.....aaagtgaagtggttctaa 1683

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues
Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :			
EST:*			
1:	em_estba.*		
2:	em_esthum.*		
3:	em_estin.*		
4:	em_estmu.*		
5:	em_estov.*		
6:	em_estpl.*		
7:	em_estro.*		
8:	em_htc.*		
9:	gb_estci.*		
10:	gb_estt2.*		
11:	gb_htc.*		
12:	gb_est3.*		
13:	gb_est4.*		
14:	gb_est5.*		
15:	em_estfun.*		
16:	em_estom.*		
17:	em_gss_hum.*		
18:	em_gss_inv.*		
19:	em_gss_pln.*		
20:	em_gss_vrt.*		
21:	em_gss_fun.*		
22:	em_gss_man.*		
23:	em_gss_mus.*		
24:	em_gss_pro.*		
25:	em_gss_rod.*		
26:	em_gss_pbg.*		
27:	em_gss_vrl.*		
28:	gb_gss1.*		
29:	gb_gss2.*		

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	848.4	50.4	1042	13	BX325239
2	767.4	45.6	1008	13	BX463023
3	698.2	41.5	891	12	BI761782
4	647	38.4	889	13	BX431866

C	5	572.8	34.0	576	14	CB156834
6	537	31.9		663	14	CB162144
7	455.8	27.1		736	14	CB594245
8	451.6	26.8		592	10	BF032018
9	447.2	26.6		802	12	BG972579
10	437.8	26.0		574	9	AA217892
11	435.4	25.9		751	12	BI148082
12	435.2	25.9		570	12	BM508620
13	416.4	24.7		929	10	BF788188
14	399.4	23.7		665	14	CF169114
15	395.4	23.6		791	12	BI332440
16	395.8	23.5		546	9	AA237499
17	392.6	23.3		608	9	AV601564
18	390.4	23.2		902	10	BF780971
19	386	22.9		682	12	BI220028
20	384.6	22.9		529	10	BF785781
21	379.8	22.6		648	14	CF171046
22	372.4	22.1		969	10	BF384535
23	370.6	22.0		654	10	AW475402
24	370.4	22.0		587	12	EG972681
25	365.6	21.7		659	10	AM610902
26	363.6	21.6		594	12	BM503097
27	363.2	21.6		625	14	CF171125
28	348.8	20.7		716	10	BF789705
29	335.8	20.0		655	10	BM569555
30	335.4	19.9		668	13	BY742633
31	332.6	19.8		733	14	CB599445
32	330.8	19.7		934	12	BI765113
33	327.8	19.5		345	14	T68666
34	323	19.2		787	14	CB955374
35	319	19.0		486	9	AA268125
36	311.8	18.5		600	10	BF788541
37	304	18.1		335	14	R89458
38	293.2	17.4		304	14	T77362
39	286.2	17.0		610	14	CA944054
40	279.8	16.6		521	12	BI221826
41	275.8	16.4		912	12	EG972626
42	275.6	16.4		366	14	T93666
43	266	15.8		434	14	T84369
44	258.6	15.4		781	10	BF533788
45	254.6	15.1		458	14	R10295

ALIGNMENTS

RESULT 1
BX325239
LOCUS BX325239 Homo sapiens T CELLS (JURKAT CELL LINE) COT 10-NORMALIZED
DEFINITION Homo sapiens cDNA clone CS000101L24 5-PRIME, mRNA sequence.
ACCESSION BX325239
VERSION BX325239.1 GI:30309195
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1042)
AUTHORS Li, W.B., Gruber, C., Jesses, J., and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 5634.f For more information about this cluster, see
http://www.genoscope.cns.fr/cgi-bin/cluster.cgi?seq=CS0AJ010DF12QP1
&cluster=5634.f Contact : Feng Liang Email : fliang@lifetech.com
URL : http://fulllength.invitrogen.com/ Invitrogen Corporation 1600 Faraday Avenue Genoscope sequence ID : CS0AJ010DF12QP1.

FEATURES	Location/Qualifiers		Db	925	AGTTACCAATGACAAAGTGAATGGGAATATGTGATGCTCTAGCTGTCTAGCCCGA	984				
	1..1042									
	/organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /clone="CS0001010Y24" /cell_type="T CELLS (JURKAT CELL LINE) COT 10-NORMALIZED" /clone_lib="Homo sapiens T CELLS (JURKAT CELL LINE) COT 10-NORMALIZED" /notes="1st strand cDNA was primed with a NotI-oligo (dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."									
source	Query Match		Qy	940	CGTTGGCTACCCAGAGAAAGCCCACTGAGCCATCAACAAGCTTCCGGGGTTTCACTC	899				
	Best Local Similarity									
	Matches 883; Conservative 6; Mismatches 7; Indels 4; Gaps 3;									
ORIGIN	10-NORMALIZED		Db	985	CGTTGGCTACCCAGAGAAAGCCCACTGAGCCATCAACAAGCTTCCGGGGTTTCACTC	1042				
	10-NORMALIZED									
	10-NORMALIZED									
ORIGIN	Query Match		Qy	985	CGTTGGCTACCCAGAGAAAGCCCACTGAGCCATCAACAAGCTTCCGGGGTTTCACTC	1042				
	Best Local Similarity									
	Matches 883; Conservative 6; Mismatches 7; Indels 4; Gaps 3;									
ORIGIN	10-NORMALIZED		Db	985	CGTTGGCTACCCAGAGAAAGCCCACTGAGCCATCAACAAGCTTCCGGGGTTTCACTC	1042				
	10-NORMALIZED									
	10-NORMALIZED									
ORIGIN	Query Match		Qy	985	CGTTGGCTACCCAGAGAAAGCCCACTGAGCCATCAACAAGCTTCCGGGGTTTCACTC	1042				
	Best Local Similarity									
	Matches 883; Conservative 6; Mismatches 7; Indels 4; Gaps 3;									
ORIGIN	10-NORMALIZED		Db	985	CGTTGGCTACCCAGAGAAAGCCCACTGAGCCATCAACAAGCTTCCGGGGTTTCACTC	1042				
	10-NORMALIZED									
	10-NORMALIZED									
ORIGIN	Query Match		Qy	985	CGTTGGCTACCCAGAGAAAGCCCACTGAGCCATCAACAAGCTTCCGGGGTTTCACTC	1042				
	Best Local Similarity									
	Matches 883; Conservative 6; Mismatches 7; Indels 4; Gaps 3;									
ORIGIN	10-NORMALIZED		Db	985	CGTTGGCTACCCAGAGAAAGCCCACTGAGCCATCAACAAGCTTCCGGGGTTTCACTC	1042				
	10-NORMALIZED									
	10-NORMALIZED									
ORIGIN	Query Match		Qy	985	CGTTGGCTACCCAGAGAAAGCCCACTGAGCCATCAACAAGCTTCCGGGGTTTCACTC	1042				
	Best Local Similarity									
	Matches 883; Conservative 6; Mismatches 7; Indels 4; Gaps 3;									
ORIGIN	10-NORMALIZED		Db	985	CGTTGGCTACCCAGAGAAAGCCCACTGAGCCATCAACAAGCTTCCGGGGTTTCACTC	1042				
	10-NORMALIZED									
	10-NORMALIZED									
ORIGIN	Query Match		Qy	985	CGTTGGCTACCCAGAGAAAGCCCACTGAGCCATCAACAAGCTTCCGGGGTTTCACTC	1042				
	Best Local Similarity									
	Matches 883; Conservative 6; Mismatches 7; Indels 4; Gaps 3;									
ORIGIN	10-NORMALIZED		Db	985	CGTTGGCTACCCAGAGAAAGCCCACTGAGCCATCAACAAGCTTCCGGGGTTTCACTC	1042				
	10-NORMALIZED									
	10-NORMALIZED									
ORIGIN	Query Match		Qy	985	CGTTGGCTACCCAGAGAAAGCCCACTGAGCCATCAACAAGCTTCCGGGGTTTCACTC	1042				
	Best Local Similarity									
	Matches 883; Conservative 6; Mismatches 7; Indels 4; Gaps 3;									
ORIGIN	10-NORMALIZED		Db	985	CGTTGGCTACCCAGAGAAAGCCCACTGAGCCATCAACAAGCTTCCGGGGTTTCACTC	1042				
	10-NORMALIZED									
	10-NORMALIZED									
ORIGIN	Query Match		Qy	985	CGTTGGCTACCCAGAGAAAGCCCACTGAGCCATCAACAAGCTTCCGGGGTTTCACTC	1042				
	Best Local Similarity									
	Matches 883; Conservative 6; Mismatches 7; Indels 4; Gaps 3;									
ORIGIN	10-NORMALIZED		Db	985	CGTTGGCTACCCAGAGAAAGCCCACTGAGCCATCAACAAGCTTCCGGGGTTTCACTC	1042				
	10-NORMALIZED									
	10-NORMALIZED									
ORIGIN	Query Match		Qy	985	CGTTGGCTACCCAGAGAAAGCCCACTGAGCCATCAACAAGCTTCCGGGGTTTCACTC	1042				
	Best Local Similarity									
	Matches 883; Conservative 6; Mismatches 7; Indels 4; Gaps 3;									
ORIGIN	10-NORMALIZED		Db	985	CGTTGGCTACCCAGAGAAAGCCCACTGAGCCATCAACAAGCTTCCGGGGTTTCACTC	1042				
	10-NORMALIZED									
	10-NORMALIZED									
ORIGIN	Query Match		Qy	985	CGTTGGCTACCCAGAGAAAGCCCACTGAGCCATCAACAAGCTTCCGGGGTTTCACTC	1042				
	Best Local Similarity									
	Matches 883; Conservative 6; Mismatches 7; Indels 4; Gaps 3;									
ORIGIN	10-NORMALIZED		Db	985	CGTTGGCTACCCAGAGAAAGCCCACTGAGCCATCAACAAGCTTCCGGGGTTTCACTC	1042				
	10-NORMALIZED									
	10-NORMALIZED									
ORIGIN	Query Match		Qy	985	CGTTGGCTACCCAGAGAAAGCCCACTGAGCCATCAACAAGCTTCCGGGGTTTCACTC	1042				
	Best Local Similarity									
	Matches 883; Conservative 6; Mismatches 7; Indels 4; Gaps 3;									
ORIGIN	10-NORMALIZED		Db	985	CGTTGGCTACCCAGAGAAAGCCCACTGAGCCATCAACAAGCTTCCGGGGTTTCACTC	1042				
	10-NORMALIZED									
	10-NORMALIZED									
ORIGIN	Query Match		Qy	985	CGTTGGCTACCCAGAGAAAGCCCACTGAGCCATCAACAAGCTTCCGGGGTTTCACTC	1042				
	Best Local Similarity									
	Matches 883; Conservative 6; Mismatches 7; Indels 4; Gaps 3;									
ORIGIN	10-NORMALIZED		Db	985	CGTTGGCTACCCAGAGAAAGCCCACTGAGCCATCAACAAGCTTCCGGGGTTTCACTC	1042				
	10-NORMALIZED									
	10-NORMALIZED									
ORIGIN	Query Match		Qy	985	CGTTGGCTACCCAGAGAAAGCCCACTGAGCCATCAACAAGCTTCCGGGGTTTCACTC	1042				
	Best Local Similarity									
	Matches 883; Conservative 6; Mismatches 7; Indels 4; Gaps 3;									
ORIGIN	10-NORMALIZED		Db	985	CGTTGGCTACCCAGAGAAAGCCCACTGAGCCATCAACAAGCTTCCGGGGTTTCACTC	1042				
	10-NORMALIZED									
	10-NORMALIZED									
ORIGIN	Query Match		Qy	985	CGTTGGCTACCCAGAGAAAGCCCACTGAGCCATCAACAAGCTTCCGGGGTTTCACTC	1042				
	Best Local Similarity									
	Matches 883; Conservative 6; Mismatches 7; Indels 4; Gaps 3;									
ORIGIN	10-NORMALIZED		Db	985	CGTTGGCTACCCAGAGAAAGCCCACTGAGCCATCAACAAGCTTCCGGGGTTTCACTC	1042				
	10-NORMALIZED									
	10-NORMALIZED									
ORIGIN	Query Match		Qy	985	CGTTGGCTACCCAGAGAAAGCCCACTGAGCCATCAACAAGCTTCCGGGGTTTCACTC	1042				
	Best Local Similarity									
	Matches 883; Conservative 6; Mismatches 7; Indels 4; Gaps 3;									
ORIGIN	10-NORMALIZED		Db	985	CGTTGGCTACCCAGAGAAAGCCCACTGAGCCATCAACAAGCTTCCGGGGTTTCACTC	1042				
	10-NORMALIZED									
	10-NORMALIZED									
ORIGIN	Query Match		Qy	985	CGTTGGCTACCCAGAGAAAGCCCACTGAGCCATCAACAAGCTTCCGGGGTTTCACTC	1042				
	Best Local Similarity									
	Matches 883; Conservative 6; Mismatches 7; Indels 4; Gaps 3;									
ORIGIN	10-NORMALIZED		Db	985	CGTTGGCTACCCAGAGAAAGCCCACTGAGCCATCAACAAGCTTCCGGGGTTTCACTC	1042				
	10-NORMALIZED									
	10-NORMALIZED									
ORIGIN	Query Match		Qy	985	CGTTGGCTACCCAGAGAAAGCCCACTGAGCCATCAACAAGCTTCCGGGGTTTCACTC	1042				
	Best Local Similarity									
	Matches 883; Conservative 6; Mismatches 7; Indels 4; Gaps 3;									
ORIGIN	10-NORMALIZED		Db	985	CGTTGGCTACCCAGAGAAAGCCCACTGAGCCATCAACAAGCTTCCGGGGTTTCACTC	1042				
	10-NORMALIZED									
	10-NORMALIZED									
ORIGIN	Query Match		Qy	985	CGTTGGCTACCCAGAGAAAGCCCACTGAGCCATCAACAAGCTTCCGGGGTTTCACTC	1042				
	Best Local Similarity									
	Matches 883; Conservative 6; Mismatches 7; Indels 4; Gaps 3;									
ORIGIN	10-NORMALIZED		Db	985	CGTTGGCTACCCAGAGAAAGCCCACTGAGCCATCAACAAGCTTCCGGGGTTTCACTC	1042				
	10-NORMALIZED									
	10-NORMALIZED									
ORIGIN	Query Match		Qy	985	CGTTGGCTACCCAGAGAAAGCCCACTGAGCCATCAACAAGCTTCCGGGGTTTCACTC	1042				
	Best Local Similarity									
	Matches 883; Conservative 6; Mismatches 7; Indels 4; Gaps 3;									
ORIGIN	10-NORMALIZED		Db	985	CGTTGGCTACCCAGAGAAAGCCCACTGAGCCATCAACAAGCTTCCGGGGTTTCACTC	1042				
	10-NORMALIZED									
	10-NORMALIZED									
ORIGIN	Query Match		Qy	985	CGTTGGCTACCCAGAGAAAGCCCACTGAGCCATCAACAAGCTTCCGGGGTTTCACTC	1042				
	Best Local Similarity									
	Matches 883; Conservative 6; Mismatches 7; Indels 4; Gaps 3;									
ORIGIN	10-NORMALIZED		Db	985						

```

Db      350 ACCCTGTGAACACGGTGGGAGTGGCTCTGTCATGGGAAACCTTCACATGACGTGCC 409
Qy      299 TGGCTCTTTCTCTGGGAATAAGTG--TCAGAAAGTGCAGAAATACGTGCAAGGACAAACC 356
Db      410 TGGCTCTTTCTCTGGGAATAAGTGATCABAHARATGCAABATACGTGCAAGGACAAACC 469
Qy      357 ATGTGGCGGGGCAATGT--CTCATACCCAGAGTCTCCCTACTACCGTGTGTCTGT 414
Db      470 ATGTGGCGGGGCAATGTATCATATACCCAGAGTCTCCCTACTACCGTGTGTCTGT 529
Qy      415 AACACACCTTACACAGGTCCAG-CTGCTCCCAAGTGGTTCCTGTATGAGGCCAAACCC 473
Db      530 HAACACCTTACACAGGTCCAGACTGCTCCCAAGTGGTTCCTGTATGAGGCCAAACCC 589
Qy      474 CTGCCAGAAATGGGGTACTCTCTCCGGCATAGCCGAGATCCAGATTCACCTGTGCTG 533
Db      590 CTGCCAGAAATGGGGTACTCTCTCCGGCATAGCCGAGATCCAGATTCACCTGTGCTG 649
Qy      534 TCCCGACCACTCAAGGGGAAATCTGTGAAATAGGTTCTGTATGACTGTATGTTGGGA 593
Db      650 TCMGACAGTCAAGGGGAAATCTGTGAATAAGTCTGTATGACTGTATGTTGGGA 709
Qy      594 TGGTACTCTTACCGAGGAAATGATAGGACAGTCAACCGAGCATGCGTGCCTTACTG 653
Db      710 TGGTACTCTTACCGAGGAAATGATAGGACAGTCAACCGAGCATGCGTGCCTTACTG 769
Qy      654 GAATCTCCACTCTCTTGCAGAGAAATACAACTATGTTATGGAGAGTCTGAAACCCA 713
Db      770 GAATCTCCACTCTCTTGCAGAGAAATACAACTATGTTATGGAGAGTCTGAAACCCA 829
Qy      714 TGGATTTGGGAAACAAATTTCTGCAGAACCCAGATCGGAGGAAAGCCCTGG-TGCT 772
Db      830 TGGATTTGGGAAACAAATTTCTGCAGAACCCAGATCGGAGGAAAGCCCTGGATGCT 889
Qy      773 TTATTAAGTTTACCAATGACAAGTGAAAT-GGGAATAGTGTGATCTCTCAGCCTGCTCA 831
Db      890 TTATTAAGTTTACCAATGACAAGTGAAATGGGAAATAGTGTGATCTCTCAGCCTACTCA 949
Qy      832 GCCCAGAGCTTGCTTACCCAGAGGAAAGCCCACTGAGCCATCAACAGTTCGGG 890
Db      950 ACCCAGACGTGCTTACCCAGAGGAAAGCCCACTGAGGATCAACCAACGTCGGG 1008

```

```

RESULT 3
BI761782 891 bp mRNA linear EST 25-SEP-2001
LOCUS 603046775F1 NIH_MGC_116 Homo sapiens cDNA clone IMAGE:5187066 5',
DEFINITION mRNA sequence.
ACCESSION BI761782
VERSION BI761782.1 GI:15753360
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cga@bbs-rcmail.nih.gov
Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L14M11467 Row: e Column: 19
High quality sequence stop: 839.
Location/Qualifiers
1..891

```

FEATURES
source

```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5187066"
/lab_host="DH108"
/clone_lib="NIH_MGC_116"
/notes="Organ: pooled colon, kidney, stomach; Vector:
pCMV-SPORT6; Site 1: NotI; Site 2: EcoRV (destroyed); RNA
source anonymous pool of 3 colons, age 26 yo male, 49 yo
female, 71 yo male colon; 46 yo male kidney, and pool of 2
stomachs, 62 yo male and 70 yo female. Library is
oligo-dT primed and directionally cloned (EcoRV site is
destroyed upon cloning). Average insert size 1.4 kb,
insert size range 1-3 kb. Library is normalized and
enriched for full-length clones and was constructed by C.
Gruber (Invitrogen). Research Genetics tracking code
023. Note: this is a NIH_MGC Library."

```

ORIGIN

```

Query Match      41.5%; Score 699.2; DB 12; Length 891;
Best Local Similarity 97.2%; Pred. No. 6.7e-188;
Matches 775; Conservative 0; Mismatches 13; Indels 9; Gaps 6;

Qy      1 ATGTTTGCAGAGATGCTCATCTCCATGTTCTGCTGTTAATGGCTCTGGTGGAAAGACA 60
Db      81 ATGTTTGCAGAGATGCTCATCTCCATGTTCTGCTGTTAATGGCTCTGGTGGAAAGACA 140
Qy      61 GCCTGTGGGTTCTCCCTGATGCTTTATTGGAAGCTGGACCCAGACTGGACCCCTGCAC 120
Db      141 GCCTGTGGGTTCTCCCTGATGCTTTATTGGAAGCTGGACCCAGACTGGACCCCTGCAC 200
Qy      121 CAGTATGATTACAGCTACGAGGATTAATATCAGGAAGAGAACACAGTAGACACTTACC 180
Db      201 CAGTATGATTACAGCTACGAGGATTAATATCAGGAAGAGAACACAGTAGACACTTACC 260
Qy      181 CATGCTGAGAAATCCTGACTGCTACTACACTGAGGACCAAGCTGATCCATGCCAGCCCCAAC 240
Db      261 CACGCTGAGAAATCCTGACTGCTACTACACTGAGGACCAAGCTGATCCATGCCAGCCCCAAC 320
Qy      241 CCCTGTGAACACGGTGGGACTGCTCTGCTCCATGAGGACACTTCACATGACGTGCTG 300
Db      321 CCCTGTGAACACGGTGGGACTGCTCTGCTCCATGAGGACACTTCACATGACGTGCTG 380
Qy      301 GCTCTCTTCTCTGGGAATAAGTGTGAGAAAGTGCAGAAATACGTGCAAGGACAAACCCATGT 360
Db      381 GCTCTCTTCTCTGGGAATAAGTGTGAGAAAGTGCAGAAATACGTGCAAGGACAAACCCATGT 440
Qy      361 GGCGGGGGCCAAATGTCATATCCAGAGTCTCTCCCTACTACCGTCTGTGCTGTAACAC 420
Db      441 GGCGGGGGCCAAATGTCATATCCAGAGTCTCTCCCTACTACCGTCTGTGCTGTAACAC 500
Qy      421 CTTTACACAGTCCAGCTGCTCCCAAGTGGTTCCTGTATGAGGCCAAACCCCTGCCAG 480
Db      501 CTTTACACAGTCCAGCTGCTCCCAAGTGGTTCCTGTATGAGGCCAAACCCCTGCCAG 560
Qy      481 AATGGGGTACCTGCTCCCGCATAGCGGAGATCCAAAGTTCACCTGTGCTGCTGCCGAC 540
Db      561 AATGGGGTACCTGCTCCCGCATAGCGGAGATCCAAAGTTCACCTGTGCTGCTGCCGAC 620
Qy      541 CAGTTCAAAGGGGAAATTCCTGCAATAGTTCCTGATGACTGCTATGTTGGCGATGGCTAC 600
Db      621 CAGTTCAAAGGGGAAATTCCTGCAATAGTTCCTGATGACTGCTATGTTGGCGATGGCTAC 680
Qy      601 TCTTACCGAGGAAATAGATAGGACAGTTC-AACAGCATGGCTGCTTTTACTGGAATC 659
Db      681 TCTTACCGAGGAAATAGATAGGACAGTTC-AACAGCATGGCTGCTTTTACTGGAATC 740
Qy      660 CC-ACCTCTCTTTCAGGAGAAATTACACATGTTTATGGAGG- --ATGCTGAAACCCCATG 715
Db      741 CCAGCTCTCTTTCAGGAGAAATTACACATGTTTATGGAGGAGTCTTGAACCCCATG 800
Qy      716 GGAATGGGAA--CACAATTTCTGAG-AACCCAGATGCGGAG-AAAAGCCCTGTGTGC 771

```

```

Db      801 GGATTGGGAAACACCAATTTCTGCAGAAAACCCAGATGCCGACGAAAAGCCCTGGTGC 860
QY      772 TTTATTAAAGTTACCAA 788
Db      861 TTTATTAAAGTACCCA 877

RESULT 4
BX431866
LOCUS   BX431866 Homo sapiens FETAL LIVER Homo sapiens cDNA clone
DEFINITION
ACCESSION BX431866
VERSION   BX431866
KEYWORDS EST.
SOURCE   EST.
ORGANISM Homo sapiens (human)

REFERENCE
AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
TITLE   Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 5634.f For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0BAG065ZH05_CS06184_1&cluster=5634.f.
Contact : Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Faraday Avenue Genoscope sequence ID : CS0BAG065ZH05_CS06184_1.
Location/Qualifiers
1. 889
source
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DM004YK02"
/tissue_type="FETAL LIVER"
/dev_stage="fetal"
/clone_lib="Homo sapiens FETAL LIVER"
/note="Organ: liver; Vector: pCMVSPORT 6; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-strand cDNA was digested with Not I and
cloned into the Not I and EcoRV sites of the pCMVSPORT 6
vector. Library was not normalized."

ORIGIN
Query Match 38.4%; Score 647; DB 13; Length 889;
Best Local Similarity 91.3%; Pred. No. 3e-173;
Matches 697; Conservative 0; Mismatches 65; Indels 1; Gaps 1;

QY      696 GGAGATGCTGAAGACCCATGGATTTGGGGAACACAATTTCTGCAGAAACCCAGATCGGA 755
Db      1 GGAGATGCTGAAGACCCATGGATTTGGGGAACACAATTTCTGCAGAAACCCAGATTCGCGA 60

QY      756 CGAAGAGCCCTGGTCTTTAATTAAGTTACCAATGACAAGTGGAATGGGAATCTGTGA 815
Db      61 CGAAGAGCCCTGGTCTTTAATTAAGTTACCAATGACAAGTGGAATGGGAATCTGTGA 120

QY      816 TGTCTCAGCCTGCTCAGCCAGGACGTTGCTTACCCAGAGAAAGCCCACTAGGCCATC 875
Db      121 TGTCTCAGCCTGCTCAGCCAGGACGTTGCTTACCCAGAGAAAGCCCACTAGGCCATC 180

QY      876 AACCAAGCTTCGGGGTTTGACTCTGTGGAAAGCACTGAGATACAGAGGAGATCAA 935
Db      181 AACCAAGCTTCGGGGTTTGACTCTGTGGAAAGCACTGAGATACAGAGGAGATCAA 240

QY      936 GAGAACTATGAGGCTTTAAGAGCACGCGGGCAAGCACCCATGGAGGCGTCCCTCCA 995

```

```

Db      241 GAGAACTATGAGGCTTTAAGAGCACGCGGGCAAGCAACCCATGGAGGCGTCCCTCCA 300
QY      996 GTCTTGCTGCTCTGACCATCTCCATGCCCCAGGGCCCACTTCTGTGGTGGGCGCTGAT 1055
Db      301 GTCTTGCTGCTCTGACCATCTCCATGCCCCAGGGCCCACTTCTGTGGTGGGCGCTGAT 360
QY      1056 CCACCCCTGCTGGGTGCTCACTGCTGCCCACTCCGACATCAAAAACAGACATCTAAA 1115
Db      361 CCACCCCTGCTGGGTGCTCACTGCTGCCCACTCCGACATCAAAAACAGACATCTAAA 420
QY      1116 GGTGGTCTTAGGGACACAGGACCTGAAGAAAGAAAGAAATTTTCATGACAGAGCTTTAGGGT 1175
Db      421 GGTGGTCTTAGGGACACAGGACCTGAAGAAAGAAAGAAATTTTCATGACAGAGCTTTAGGGT 480
QY      1176 GGAGAAGATATTCAAGTACAGCCACATACATGAAGAGATGAGATTTCCCAATGATAT 1235
Db      481 GGAGAAGATATTCAAGTACAGCCACATACATGAAGAGATGAGATTTCCCAATGATAT 540
QY      1236 TGCATTCTCAAGTTAAAGCCAGTGGATGCTCACTGTGCTCTAGAAATCCAAATACGTGAA 1295
Db      541 TGCATTCTCAAGTTAAAGCCAGGATGGTACTGTGCTCTAGAAATGCAAAATACGGGAA 600
QY      1296 GACTGTGCTTGCTGCTGATGGGTCCCTTCCCTCTGGAGTGAAGTCCACAT-CTCTGGCT 1354
Db      601 GACTGTGGGCTTGCTGCTGATGGGTCCCTTCCCTCTGTAGAGAGTGCCACATGTTGTGACT 660
QY      1355 GGGGTGTTACAGAAACAGGAAAGGGTCCGCGCAGCTCCTGGATGCCAAAGTCAAGCTGA 1414
Db      661 GTGGGGTTCAGAAACAGTTAAAGTTTCCCTCCGCTACTGTTCCCTAAGATAGTTTCG 720
QY      1415 TTGCAACACATTTTGCAACTCCCGCAACTCTATGACACAT 1457
Db      721 TTGTCGACATGGAGGGGGGGGGTGGTGGTACTGTGTGTGCACAT 763

RESULT 5
BX156834/c
LOCUS   BX156834 K-EST0215807 L17N670205n1 Homo sapiens cDNA clone
DEFINITION
ACCESSION BX156834
VERSION   BX156834
KEYWORDS EST.
SOURCE   EST.
ORGANISM Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 576)
Kim, N.S., Hahn, Y., Oh, J.H., Lee, J.Y., Ahn, H.Y., Chu, M.Y., Kim, M.R.,
Oh, K.J., Cheong, J.E., Sohn, H.Y., Kim, J.M., Park, H.S., Kim, S. and
Kim, Y.S.
21C Frontier Korean EST Project 2001
Unpublished (2002)
Contact: Kim YS
Genome Research Center
Korea Research Institute of Bioscience & Biotechnology
52 Eoeun-dong yuseong-gu, Daejeon 305-333, South Korea
Tel: +82-42-860-4470
Fax: +82-42-860-4409
Email: yongsung@mail.kribb.re.kr
Plate: 4 row: F column: 07
High quality sequence stop: 576.
Location/Qualifiers
1. 576
source
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="L17N670205n1-4-F07"
/sex="F"
/lab_host="Top10P"
/clone_lib="L17N670205n1"
/note="Organ: Liver; Vector: p7T73-Pac; Site 1: EcoRI;
Site 2: NotI; The library was contributed by the Soares

```

ORIGIN

Query Match	34.08;	Score 572.8;	DB 14;	Length 576;
Best Local Similarity	99.7%;	Prod. No. 3.8e-152;		
Matches 574;	Conservative 0;	Mismatches 152;	Indels 0;	Gaps 0;
494	QY	GCTCCCGGCATAAGCGGAGATCCAAGTTTCACTGTGCCCTGTCCGACCAAGTTCGAAGGGA	553	
576	DB	GCTCCCGGCATAAGCGGAGATCCAAGTTTCACTGTGCCCTGTCCGACCAAGTTCGAAGGGA	517	
554	QY	AATTCGTGTGAATAGGTTCTGTAGACTGTCTATGCTTGGCGATGGCTACTCTTACCGAGGA	613	
516	DB	AATTCGTGTGAATAGGTTCTGTAGACTGTCTATGCTTGGCGATGGCTACTCTTACCGAGGA	457	
614	QY	AAATGAATPAGACAGTCAACACAGATGCGTGCCCTTTACTGTGGAACCTCCCACTCTCTTTCG	673	
456	DB	AAATGAATPAGACAGTCAACACAGATGCGTGCCCTTTACTGTGGAACCTCCCACTCTCTTTCG	397	
674	QY	AGGAGAAATTCAAACATGTTTATGGAGATGCTGAAACCCATGGGATGGGGAAACAATTT	733	
396	DB	AGGAGAAATTCAAACATGTTTATGGAGATGCTGAAACCCATGGGATGGGGAAACAATTT	337	
734	QY	TCTCAGAAAACCCAGATGCGGACGAAAGCCCTGTGTCTTTATTAAGTTTCCATGTACA	793	
336	DB	TCTCAGAAAACCCAGATGCGGACGAAAGCCCTGTGTCTTTATTAAGTTTCCATGTACA	277	
794	QY	AGGTGAATGGGAATCTGTGATCTCTCAGCCTGCTCAGCCAGGACGTTGCTTACCAG	853	
276	DB	AGGTGAATGGGAATCTGTGATCTCTCAGCCTGCTCAGCCAGGACGTTGCTTACCAG	217	
854	QY	AGGAAGCCCACTGAGGCCATCAACCAAGCTTCGGGGTTTGACTCCTCTGGAAGACGTG	913	
216	DB	AGGAAGCCCACTGAGGCCATCAACCAAGCTTCGGGGTTTGACTCCTCTGGAAGACGTG	157	
914	QY	AGATAGCAGAGAGGAAGATCAAGAGAATCTATGAGGCTTTAAGACGACGGCGGGCAAGC	973	
156	DB	AGATAGCAGAGAGGAAGATCAAGAGAATCTATGAGGCTTTAAGACGACGGCGGGCAAGC	97	
974	QY	ACCCATGCGAGCGTCCCTCCAGTCTCTCGTGTCTGACCACTCCATGCCCCCAGGGCC	1033	
96	DB	ACCCATGCGAGCGTCCCTCCAGTCTCTCGTGTCTGACCACTCTCATGCCCCCAGGGCC	37	
1034	QY	ACTTCTGTGTGGGCGCTGATCCACCCCTGCTGGG	1069	
36	DB	ACTTCTGTGTGGGCACTGATCCACCCCTGCTGGG	1	

RESULT 6	CB162144	663 bp	mRNA	linear	EST 30-JAN-2003
LOCUS	CB162144				
DEFINITION	K-EST0222650 LI7N670205n1 Homo sapiens cDNA clone				
	LI7N670205n1-17-H10 5', mRNA sequence.				
ACCESSION	CB162144				
VERSION	CB162144.1	GI:28148270			
KEYWORDS	EST.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
REFERENCE	1. (bases 1 to 663)				
AUTHORS	Kim,N.S., Hann,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R., Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and Kim,Y.S.				
TITLE	21C Frontier Korean EST Project 2001				
JOURNAL	Unpublished (2002)				
COMMENT	Contact: Kim YS Genome Research Center Korea Research Institute of Bioscience & Biotechnology 52 Eoeun-dong Yuseong-qu, Daejeon 305-333, South Korea				

Tel: +82-42-860-4470
Fax: +82-42-860-4409
Email: yongsung@mail.kribb.re.kr
Plate: 17 row: H column: 10
High quality sequence stop: 863.

FEATURES
SOURCE

```

source
1. .663
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="L17N670205n1-17-H10"
/sex="F"
/lab_host="Topi0F"
/clone_lib="L17N670205n1"
/name="Torgan: Liver: Vector:
Site_2: NoI; The library was
laboratory and it was constr
M.F., Lennon, G. and Soares,
6(9): 791-806. RNA was prepa
culture."

```

ORIGIN

Query Match	31.9%;	Score 537;	DB 14;	Length 663;
Best Local Similarity	100.0%;	Pred. No. 7.3e-142;		
Matches 537;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1147	GAAGAAATTCATGACAGACAGCTTTAGGGTGGAGAGATATTCAAGTACAGCCACTACAAT	1206	
DB	1	GAAGAAATTCATGACAGACAGCTTTAGGGTGGAGAGATATTCAAGTACAGCCACTACAAT	60	
QY	1207	GAAGAGATGAGATTCCTCCACCAATGATATTGCAATTCCTCAAGTTAAAGCCAGTGGATGTT	1266	
DB	61	GAAGAGATGAGATTCCTCCACCAATGATATTGCAATTCCTCAAGTTAAAGCCAGTGGATGTT	120	
QY	1267	CACTGTGCTCTAGAAATCCAAATACGTAAGACATGTGTGCTTGCTTGCCTGATGGGTCTCTTCCC	1326	
DB	121	CACTGTGCTCTAGAAATCCAAATACGTAAGACATGTGTGCTTGCTTGCCTGATGGGTCTCTTCCC	180	
QY	1327	TCCTGGAGTGGTGCACATCTCTGCTGGGTGTTTACAGAAACAGAAAAAGGGTCCCGC	1386	
DB	181	TCCTGGAGTGGTGCACATCTCTGCTGGGTGTTTACAGAAACAGAAAAAGGGTCCCGC	240	
QY	1387	CAGCTCCTGGATGCCAAAGTCAAGCTGATTGGCCAAACATTTGTGCAACTCCGCGCAACTC	1446	
DB	241	CAGCTCCTGGATGCCAAAGTCAAGCTGATTGGCCAAACATTTGTGCAACTCCGCGCAACTC	300	
QY	1447	TATGACACATGATTGATGACAGATATGATCTGTGCAGGAAATCTTCAGAAAACTGGGCAA	1506	
DB	301	TATGACACATGATTGATGACAGATATGATCTGTGCAGGAAATCTTCAGAAAACTGGGCAA	360	
QY	1507	GACACCTGCCAGGTCACATCTCGAGGCCCTGACCTGTGAGAAAGCAGGCACCTACTAC	1566	
DB	361	GACACCTGCCAGGTCACATCTCGAGGCCCTGACCTGTGAGAAAGCAGGCACCTACTAC	420	
QY	1567	GTCATATGGGATAGTGAAGTGGGGCTGGAGTGGGAAAGAGGCCAGGGGTCTACCCCAA	1626	
DB	421	GTCATATGGGATAGTGAAGTGGGGCTGGAGTGGGAAAGAGGCCAGGGGTCTACCCCAA	480	
QY	1627	GTTACCAAATTCCTGAATTTGGATCAAAGCCACCATCAAAGTCAAAAGTGGCTTCTAA	1683	
DB	481	GTTACCAAATTCCTGAATTTGGATCAAAGCCACCATCAAAGTCAAAAGTGGCTTCTAA	537	

RESULT 7

AB031177	CB594245	CB594245	736 bp	trna	linear	EST 03-APR-2003
LOCUS	CB594245	AGENCOURT_12929103	NH_MGC_177	Mus musculus	CDNA clone	
DEFINITION		IMAGE:30310462 5', mRNA sequence.				
ACCESSION		CB594245				
VERSION		CB594245.1	GI:29512101			
KEYWORDS		EST.				
SOURCE		Mus musculus (house mouse)				
ORGANISM		Mus musculus				

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 736)
NIH-MGC <http://mgs.nci.nih.gov/>
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: Dr. Michael Brownstein
cDNA Library Preparation: Michael Brownstein Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Plate: NDCM92 row: d column: 23
High quality sequence stop: 558.
Location/Qualifiers
1..736
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:30310462"
/lab_host="DH10B (T1-phage-resistant)"
/clone_lib="NIH MGC 177"
/note="Organ: liver; Vector: pDNR-LIB; Site: 1: SfiI (ggcattagggcc); Site 2: SfiI (ggcgctgggccc); cDNA made by oligo-dT priming and directionally cloned. 5' and 3' adaptors were used in cloning as follows:
5'-AAGCAGTGTATCAGCAGGAGCGCATTCAGCGCGG-3' and 5'-ATTCTAGACCGGCGGCGGACATG-AT(30)NN-3'. Full-length enriched library was constructed using the Clontech Creator SMART kit and size-selected to contain the 0.5 kb size fraction. Library created in the laboratory of M. Brownstein (NIH, NIH). Note: this is a NIH_MGC Library."

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

FEATURES
source

ORIGIN

Query Match 27.1%; Score 455.8; DB 14; Length 736;
Best Local Similarity 80.9%; Pred. No. 1.3e-118;
Matches 543; Conservative 0; Mismatches 127; Indels 1; Gaps 1;

Qy 715 GGGATTGGGAACACATTTCTGCAGAACCCAGATCGGACGAAAGCCCTGGTCTTT 774
Db 3 GGGATCGCAGAGCAACATCTCTCGAGAACCCAGATGGAGACCAACCCCTGGTCTT 62
Qy 775 ATTAAGATTACCAATGACAAAGTGGAAATGGAAATCTGTATCTCTCAGCCTCTCAGCC 834
Db 63 GTCAGGTGACAGTGAAGAGTGAATGGAAATCTGTATCTCAGTCTCTCAGTG 122
Qy 835 CAGGACCTGTGCTTACCCAGAGGAAAGCCCACTGAGCCATCAACCAAGCTTCGGGGTTT 894
Db 123 CTTGACACCCCTACCCAGTGGAAAGCCCTTCTGAGCCTGTGATGGAGCTGCCAGGGTTC 182
Qy 895 GACTCTGTGAAAGACTGAGATGACAGAGAGGAAAGATCAAGAAATCTATGAGGCTTT 954
Db 183 GAGTCTTGGGGAAGAGAGGATGATGACACGCACTCAAGCGTATCTACGGGGCTTT 242
Qy 955 AAGAGCAGCGCGGCAAGCAACCCATGCGAGCGTCCCTCCAGTCTCTGCTGCTCTGACC 1014
Db 243 AAGAGCAGCAGGCAAGCAACCCGTGCGAGTGTCTCTGCGAGACCTCACTGCGGTGACC 302
Qy 1015 ATCTCCATGCCAGGCGCACTTCTGTGGGCGCTGATCCACCCCTCTGGTGGTCTC 1074
Db 303 ACCTCCATGCCAGGCGCACTTCTGTGGGCGCTGATCCACCCCTCTGGTGGTCTC 362
Qy 1075 ACTGCTGCCACTGACCGACATAAAAACAGACATCTAAAGTGGTGTCTAGGGACCG 1134
Db 363 ACTGACGCCACTGTACCGACATAAACCAACCAAGCATCTAAAGTTGTACTAGGGATCAG 422
Qy 1135 GACCTGAGAAAGAGATTTTCATGACGAGCTTTAGGCTGGAGAGATTCAGATAC 1194
Db 423 GACCTGAGAAAGACAGAAATCCATGAACACCTTCAGGGTGGAAAAAATCTGAAGTAC 482

Qy 1195 AGCCACTACATGAAGAGATGAGATTCGCCCAATGATATTCATGCTCAAGTTAAAG 1254
Db 483 AGTCAGTATATGAAGAGATGAGATTCGCCCAATGATATTCATGCTCAAGTTAAAG 542
Qy 1255 CCAGTGGATGGTCACTGTGCTCTAGAAATCCAAATAGTGAAGACTGTGCTTGCCTGAT 1314
Db 543 CCAGTGGTGGTCACTGTGCTCTGGAATCCAGATATGTGAGACTGTATGTTGCCAGC 602
Qy 1315 GGGTCTTTCCCTCTGGGAGTGGAGTGCACATCTCTGGCT-GGGGTGTTACAGAAACAGG 1373
Db 603 GACCCCTTTCCCTCTGGAACTGAGTGCACATCTCTGGCTGGGGGTGTTTCAGAAACAGG 662
Qy 1374 AAAAGGGTCCC 1384
Db 663 GGAAGGGGTCC 673

RESULT 8
BE032018
LOCUS BE032018 592 bp mRNA linear EST 09-JUL-2000
DEFINITION 130809 MARC 1P1G Sus scrofa cDNA 5', mRNA sequence.
ACCESSION BE032018
VERSION BE032018.1 GI:8327027
KEYWORDS EST.
SOURCE Sus scrofa (pig)
ORGANISM Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
1 (bases 1 to 592)
Fahrenkrug, S.C., Smith, T.P.L., Preking, B.A., Cho, J., White, J., Vallet, J., Wise, T., Rohrer, G.A., Pertea, G., Sultana, R., Quackenbush, J. and Keeler, J.W.
Porcine gene discovery by normalized cDNA-library sequencing and EST cluster assembly
Mamm. Genome 13 (8), 475-478 (2002)
22213789
12226715

COMMENT
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smithemail.marc.usda.gov
Single pass sequencing. Bases called and alt trimmed with phred v0.980904.e. Vector identified by cross_match with the -minscore 18 and -minmatch 12 options.
PCR primers
FORWARD: AGGAAACAGCTATGACCAT
BACKWARD: GTTTCCTCAGTCACGAC
Plate: 69 row: K column: 12
Seq primer: ATTAGTGACACTATAG.
Location/Qualifiers
1..592
/organism="Sus scrofa"
/mol_type="mRNA"
/db_xref="taxon:9823"
/tissue_type="pooled"
/lab_host="DH10B"
/clone_lib="MARC 1P1G"
/note="Vector: pCMV SPORT6; Site: 1: NotI; Site 2: SalI; Library made from pooled tissue from day 11, 13, 15, 20, and 30 embryos."

FEATURES
source

ORIGIN

Query Match 26.8%; Score 451.6; DB 10; Length 592;
Best Local Similarity 86.0%; Pred. No. 1.8e-117;
Matches 518; Conservative 0; Mismatches 69; Indels 15; Gaps 1;

Qy 522 CACCTGTGCTCTCCCGACCACTTCAAGGGGAAATCTGTGAAATAGGTTCTGATGACTG 581
Db 6 CGCTGGGCGCTCTCTGACCACTTAAAGGAGATCTGTGAATAGGTTCTGATGACTG 65
Qy 582 CTATGTTGGCAGTGGCTACTCTTACCAGGGGAAATGATAGGACATCACCAGCATGC 641


```
Db 66 CTATGTTGGTACCGCTACTCTTACCGAGGGAAGTGAATAAACTGTCAACGACAC 125
Qy 642 GTGCTTTTACTGGAACCTCCACTCTCTTTCAGGAGAAATPACAACTGTTATGGAGGA 701
Db 126 GTGCTTTTACTGGAACCTCCACTCTCTTTCAGGAGAAATPACAACTGTTATGGAGGA 185
Qy 702 TGCTGAACCCATGGATTTGGGAACCAATTTCTGCAGAAACCCAGATCGGACGAAA 761
Db 186 TGCTGAGCCCATGGATTTGGGAGCACACTTCTGCAGAAACCCAGATGGAGAGAAA 245
Qy 762 GCCCTGTGTCTTTATTAAGTTACCAATGACAAGGTGAAATGGAATACCTGATGTCTC 821
Db 246 GCCCTGTGTCTTTATTAAGTTACCAATGACAAGGTGAAATGGAATACCTGATGTCTC 305
Qy 822 AGCTGCTCAGCCAGGAGCTTGCCTACCCAGAGGAAAGCCCACTGAGCCATCAACAA 881
Db 306 TGCTGTCTCAGC-----CCGAGAGGGAAGCCCACTGAGAACTGTGACCGA 350
Qy 882 GCTTCCGGGGTTTGACTCCTGTGGAAGACTGAGATAGCAGAGAGGAAGATCAAGAAAT 941
Db 351 GCTTCCCGGGTTTGACTCCTGTGGGAGCAGACAGCAGAAAGGAAGTTCAAGAGAT 410
Qy 942 CTATGGAGGCTTTAAGACAGCGGGGCAAGCAACCATGAGGAGGCTCCCTCAGTCTC 1001
Db 411 CTACGAGGCTTTAAGACAGCGGGGCAAGCAACCATGAGGAGGCTCCCTACAGACCTC 470
Qy 1002 GCTGCTCTGACCACTCTCATGCCCGGCGCACTTCTGTGGGGGCTGTATCCACCC 1061
Db 471 CTGCAATCTGACCGTCTCATGCCCGGCGCACTTCTGTGGGGGCGCTGATCCACCC 530
Qy 1062 CTGCTGGGTGCTCACTGTGCGCACTGCAACCGACATAAAACCCAGCATCTAAAGGTGGT 1121
Db 531 CTGCTGGGTGCTCACTGTGCGCACTGCAACCGACATAAAAGCCAAATATCTAAAGTAGT 590
Qy 1122 GC 1123
Db 591 GC 592

RESULT 9
Bg972579 802 bp mRNA linear EST 12-JUN-2001
LOCUS 602841165F1 NCI_CGAP_Kid14 Mus musculus cDNA clone IMAGE:4975384
DEFINITION 5', mRNA sequence.
ACCESSION Bg972579
VERSION Bg972579.1 GI:14360216
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
NIH-MGC http://mgi.ncl.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L14M10967 row: a column: 17
High quality sequence stop: 785.
Location/Qualifiers
1..802
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N"
/db_xref="taxon:10090"
```

FEATURES

source

RESULT 10
AA217892
LOCUS
DEFINITION mv54b08.r1 Soares mouse 3NBE12 5 Mus musculus cDNA clone
IMAGE:658839 5, similar to TR:G1345398 G1345398 HGF ACTIVATOR LIKE
PROTEIN. ;, mRNA sequence.
ACCESSION AA217892

/clone="IMAGE:4975384"
/lab_host="DH10B (T1 phage-resistant)"
/clone_lib="NCI_CGAP_Kid14"
/note="Organ: Kidney; Vector: pCMV-SPORT6; Site 1: Nof1;
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dr.
Average insert size 1.75 kb. Constructed by Life
Technologies. Note: this is a NCI_CGAP Library. |"

ORIGIN

Query Match 26.6%; Score 447.2; DB 12; Length 802;
Best Local Similarity 77.7%; Pred. No. 4e-116;
Matches 580; Conservative 0; Mismatches 158; Indels 8; Gaps 3;

Qy 1 ATGTTTGCAGAGATGTCGATCTCCATGTTCTGCTGTTAATGGCTCTGGTGGAAAGACA 60
Db 50 AATATTGTGAGATGTTGGTGTTCGCTGCTGCTAATCGCTGCTGGTGGGAAGTCA 109
Qy 61 GCTGTGGGTTCTCCCTGATGTCCTTATGGAAGCTGGACCCAGACTGGACCCCTGAC 120
Db 110 GTCATTGGGCTCTCACTGATGTCCTTATGCGCCCGCCAGACCCAGATTGGACCCCGAT 169
Qy 121 CAGTATGATTACAGCTACGAGGATTATATCAGGAAGAGAACACCACTAGTAGCACACTTACC 180
Db 170 GACTATTACTACAGCTATGAGCAGTCCAGCCAGACGAGACCCCGAGTGTACGAGACC 229
Qy 181 CATGTGGAATCTGACTGTGTTACTACTGAGGACCAAGCTGATCCATGCGAGCCCAAC 240
Db 230 ACCCTTGAGAACCCCGACTGCTACT-----ATGAGACGATGATCCATGCGAGTCCAAC 283
Qy 241 CCTGTGAACAGGTGGGAGCTGCTGCTCCATGGAGACCTTCACATGACAGCTGCCCTG 300
Db 284 CCTGTGAACAGCGCGGGACCTGATCATCAGAGGGGATACCTTCAGTTGCACTGCCCCA 343
Qy 301 GTCCTCTTCTCTGGGAATTAAGTGTGAGAAAGTGCAGAAATACGTGCAAGGACCAACCATGT 360
Db 344 GCCCTCTTCTCGGGAGCGGTGCCAGACTGCCACAAAACAAGTGAAGGACCAACCATGT 403
Qy 361 GCGCGGGCAATGTCCTATCCAGAGTCTCCCTACTACCGCTGCTCTGTGTAACAC 420
Db 404 GTCCTGTTGATGTCCTCATTTACCCAGAGACCCCTACTACCGCTGCTCTGTGTAACAC 463
Qy 421 CCTTACACAGGTCCCAAGCTGCTCCCAAGTGTTCCTGTATGCAAGGCAACCCCTGCCAG 480
Db 464 CCTTACACGGGACCAAGCTGCTCCCAAGTGTTCCTGCGCATGCAAGGCAACCCCTGCCAG 523
Qy 481 AATGGGGTACCTGCTCCCGGATAGCGGAGATCCAGTTCACCTGTGCTGCTCCCGAC 540
Db 524 AATGGCGGAGTCTGTTCCCGACACAGCGAGATCCAGTTTACCTGTGCTGCTCCAGAC 583
Qy 541 CAGTTCAGGGGAAATTTCTGTGAA-ATAGGTCTGATGACTGCTATGTTGGGATGGCTA 599
Db 584 CAGTATAAGGGGAAATTTCTGTGAACATAGTCCGACGACTGTTATGTCGTTGATGGCTA 643
Qy 600 CTCTTACCGAGGAAATGAATAGGACAGTCAACGAGCATGCGTTCCTTACTGGAACTC 659
Db 644 CTCTTACCGAGGAAATGATGAAGACAGTCAACCAAGACCCAGGSCCTTTACTGGAACT- 702
Qy 660 CCACCTCTCTTCGAGGAGAAATTAACAATGTTTATGGAGGATGCTGAAACCCATGGGAT 719
Db 703 CCACTCTCTTCGAGGAGACTTATACATGTTTATGGAGGATGCAAGACCCACCGGAT 762
Qy 720 TGGGGAACACAATTTCTGAGAAACC 745
Db 763 GCGAGAGACAACTTCTGAGAAACC 788

	Query Match	26.0%	Score 437.8;	DB 9;	Length 674;
	Best Local Similarity	78.2%;	Pred. No. 1.8e-113;		
	Matches 526;	Conservative 0;	Mismatches 147;	Indels 0;	Gaps 0;
QY	701	ATGCTGAACCCCATGGATTGGGGAACACAAATTTCTGCAGAAACCCAGATGCGGACGAA	760		
Db	1	ATTCCGGAAGAGGCACGGATCGCAGACACAACTTCTGCAGAAACCCAGATGGAGACCA	60		
QY	761	AGCCCTGCTCTTATTAAAGTTACCAATACACAGGTGGAATGGGAATACCTGTGATGTCT	820		
Db	61	AACCCCTGGTTCGTCAAGTGAACAGTGAGAGGTGGAATGGGAATCTGTGATGTCA	120		
QY	821	CAGCCTCTCTCAGCCACGAGCGTTGCTCTACCCAGAGGAAACCCCACTGAGCCATCAACCA	880		
Db	121	CAGTCTCTCAGTGCCTTGACACCCCTTAACCCAGTGGAAAGCCCTTCTGGAGCCTGTGTATGG	180		
QY	881	AGCTTCGGGGTTTGACTCTCTGTGGAAAGACTGAGATGACAGAGGAGAAAGTCAAGAGAA	940		
Db	181	AGCTGCCAGGGTTTCAGTCTCTGCGGAAGACGGAAGTAGTGTACACGCACTCAAGGTA	240		
QY	941	TCTATGAGAGCTTTTAAGAGCACGCGGGCAAGCAACCCATGCGAGGCTCTCCCTCCAGTCCT	1000		

Query Match	25.9%	Score 435.4	DB 12	Length 751
Best Local Similarity	78.0%	Prsd. No. 9.2e-113		
Matches 589	Conservative 0	Mismatches 156	Indels 10	Gaps 5
QY	4	TTTGCCAGAGTGTCTGATCTCCATGTTCTGCTGCTTAATGGCTCTGCTGGAGAAACACACG- 62		
Db	2	TTTGTCCAGGATGTTGGTGTTCCTGGTCTCTCTTAATCGCCCTGGTGGGNACTCAGTG 61		
QY	53	CTGTGGGTTCTCCCTGATGTCTTTATTGGAAAGCCCTGGACCCAGACTGGACCCCTTGACCA 122		
Db	62	CATTGGGCTCTCACTGATGTGCTTTCATTTGGCCGCCCCAGACCCAGATTGGACCCCGCATGA 121		
QY	123	GTATGATTACAGCTACGAGGATATTAATCAGGAAGAGAAACACAGTAGCACACTTACCACA 182		
Db	122	CTATTACTACAGCTATGAGCAGTCCAGCCCCAGACGAAGACCCAGCTGTCCAGCAGCCAC 181		
QY	193	TGCTGAGAAATCCTGACTGTGTACTACTGAGGACCAAGCTGATTCATGCCAGCCCAACCC 242		
Db	182	CCCTGAGAACCCCGACTGGTACT-----ATGGAAGCATGATTCATGCCAGTCCAAACCC 235		
QY	243	CTGTGAACACAGGTGGGAGCTGCCTCTGTCCTACATGGGAGCACCTTCACATCAGCTGCCTGGC 302		
Db	236	CTGTGAACACAGCGGGGACTGTATCATCAGAGGGGATACCTTCAGTTGCAGTGCCTCCAGC 295		
QY	303	TCCCTTTCCTGGGAATTAAGTGTGAAAGTGCAAATATAGTGCNAGGACCAACCCATGTGG 362		
Db	296	CCCCCTTCCTGGGAGACCGGTGCAGACTGCACAAAACAAAGTCAAGGCAACCCATGTGTT 355		
QY	363	CCGGGCCCAATGTCTCATTTACCAGAGTCTCCTCTACTACCGCTGTGTCTGTAAACACCC 422		
Db	356	CCATGGTGATTGTCCTCATTTACCCAGAGCACCCCTACTACCGCTGTGCTGCAATATACC 415		
QY	423	TTACACAGGTCCTGAGCTGCTCCGAAAGTGTTCCTGTATGCAGGCCCAACCCCTGCCAGAA 482		
Db	416	TTACACGGGACACAGCTGCTCAAAGTGTCTCGGCATGCAGGCCAAACCCCTGCCAGAA 475		
QY	483	TGGGCTACCTGCTCCCGGCATAGCGGAGATCCAAAGTTCACTGTGCTGTCTCCCGACCA 542		
Db	476	TGGCGGAGTCTGTTCCCGACACACAGCGGAGATCCAGGTTTACTGTGCTGTCTCCAGACCA 535		
QY	543	GTTCAAGGGGAAATCTGTGAAATAGGTTCTGATGTACTGTATGTGGCGATGGCTACTC 602		
Db	536	GTTAAGGGGAAATCTGTGAAATAGTTCGGGACGACTGTTATGTGCGGTGATGGGTACTC 595		
QY	603	TTACCGAGGGAAAATGATTAGGACAGTCAACACGACATGCGTGCTCTTTACTGG-AACTCCC 661		
Db	596	TTACCGAGGCAAGTGAGTAAGA CAGTCAACGAGAACCCATGCTTTACTGGAAACTCCC 655		
QY	662	ACCTCCTCTTCGAGGAGNAATTACACATGTTTATGAGGATGCTGAACCCATGGGATTG 721		
Db	656	ACCTCCTCTTCAGAGACCTTATTAACATGTTTATGAGGATGCAGAGACCCACGGGA-TC 714		
QY	722	GGGAACAAATTTCTGCAGAAACCCAGATCGGAC 756		
Db	715	GAGAGCAAACTTTGTG-AGAAACCCAGATCGAGAC 748		

RESULT 12	BM508620	570 bp	linear	EST 12-MAR-2002
LOCUS	BM508620			
DEFINITION	.ii37h03.yl Melton Normalized Mixed Mouse Pancreas 1 NI-MMS1 Mus			
	HGF ACTIVATOR LIKE PROTEIN. ;, similar to TR:Q14520 Q14520			
ACCESSION	BM508620			
VERSION	BM508620.1	GI:18679763		
KEYWORDS	EST			
SOURCE	Mus musculus (house mouse)			
ORGANISM	Mus musculus			
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
REFERENCE	1 (bases 1 to 570)			
AUTHORS	Leiton,D., Brown,J., Keny,G., Permut,A., Lee,C., Kaesner,K.,			
	Menishka,I., Searce,M., Brestelli,J., Gadowh,J.G., Clifton,S.,			

Hillier, L., Marra, M., Pape, D., Wylie, T., Martin, J., Blistain, A.,
 Schmitt, A., Thesling, B., Ritter, E., Ronko, I., Bennett, J.,
 Cardenas, M., Gibbons, M., McCann, R., Cole, R., Tsagarisvili, R.,
 Williams, T., Jackson, Y., and Bowers, Y.
 Endocrine Pancreas Consortium
 Unpublished (2000)
 Other ESTs: i137h03.x1
 Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
 Endocrine Pancreas Consortium
 Harvard University, Howard Hughes Medical Institute
 Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
 MA 02138
 Tel: 617-495-1812
 Fax: 617-495-8557
 Email: dmelton@iobhp.harvard.edu
 Library was constructed by Dr. Douglas Melton DNA sequencing by:
 Washington University Genome Sequencing Center For information on
 obtaining a clone please contact: Juliana Brown
 (brownja@wustl.edu)
 MGI:2007028 This sequence now available from the IMAGE consortium,
 for clone orders contact: info@image.llnl.gov
 Seq primer: -40RP from Gibco
 High quality sequence stop: 435.

[illegible]

```

1. .570
Location/Quailires
/organism="Mus musculus"
/mol_type="mRNA"
/strain="ICR"
/db_xref="taxon:10090"
/clone="IMAGE:5944156"
/sex="Both for embryonic & newborn, male for adult and
adult islet"
/dev_stage="Embryonic day 10.5, E12.5, E16.5, newborn,
adult, mixed"
/lab_host="DH103"
/clone_lib="Melton Normalized Mixed Mouse Pancreas 1
NI-MMS1"
/notes="Vector: pSPORT1; Site_1: Not I; Site_2: Sal I; Five
libraries representing E10.5/12.5 pancreatic bud, E16.5
pancreas, newborn pancreas, adult pancreas, and adult
islets of Langerhans were separately constructed using
SuperScript Plasmid Library kit (Life Technologies). cDNA
was made by oligo-dT priming and size-selected by column
fractionation. Libraries were amplified once on solid
support and plasmid DNA from each library was prepared
and mixed in equal amounts. The mixed library DNA was
normalized by method #4 from Bonaldo, Lennon, and Soares
1996 Genome Research 6:791-806; 0.5 microgram
single-stranded mixed library plasmid DNA was mixed with
5 micrograms PCR product representing mixed library
inserts and hybridized to an EcoT of 6. Single-stranded
(unhybridized) plasmids were isolated by hydroxyapatite
chromatography and used to make this library.

```

ORIGIN

	Query Match	25.98;	Score 435.2;	DB 12;	Length 570;
	Best Local Similarity	85.4%;	Pred. No. 8.8e-113;		
	Matches 485;	Conservative 0;	Mismatches 83;	Indels 0;	Gaps 0;
Qy	1108	CATCTAAAGGTGGTCTAGGGGACCGAGGACCTCTGAAGAAGAGAAATTTTCATGACGACAGC	1167		
Db	3	CATCTAAAGGTGTACTAGGGGATCAGGACCTGGAAGAAGACAGAAATCCCATGAACAGACC	62		
Qy	1168	TTTAGGGTGGAGAGATATTCAAGTACAGCCACTACAATGAAGAGATGAGATTTCCCCAC	1227		
Db	63	TTCAGGTTGGAAAATACTGAAGTACACTCAGTATATATGAAGAAGATGAGATTTCCCCAC	122		
Qy	1228	AATGATATTGCAATTGCTCAAGTTAAAGCCAGTGGAGTGGTCACTGTGCTCTAGATCCAAA	1287		
Db	123	AATGACATGTGTTTGCCTCAAGTTAAAGCCAGTGGGTGGTCACTGTGCTCTGGAATCCAGA	182		
Qy	1288	TACGTGAAGACTGTGTGCTTTGCGCTGATGGGTCTCTTCCCTCTGGGAGTGAAGTGCACATC	1347		

Db 183 TATGTGAAGACTGTATGTTGTCAGGACCCCTTTCTGGAACCTGAAGTGCACATC 242
Qy 1348 TCTGGCTGGGTGTTTACAGAAACGGAAGAGGTCCCGCAGCTCCTGGATGCCAAAGTC 1407
Db 243 TCTGGCTGGGTGTTTACAGAAACGGAAGAGGTCCCGCAGCTCCTGGATGCTAAAGTC 302
Qy 1408 AAGCTGATTGCCAACACATTTGTGCAACTCCCGCAACTCTATGACCAATGATTGATGAC 1467
Db 303 AAGCTAATCGTAAACCTTTGTGCAACTCCCGCAACTCTATGACCAATGATTGATGAC 362
Qy 1468 AGTATGATCTGTGCAGGAATCTTCAGAAACCTGGCAAGACACCTGCCGGTGACTCT 1527
Db 363 AGTATGATTGTGGGGGAACCTTCAGAAAGCCGGATCAGACACCTGCCAGGTGACTCG 422
Qy 1528 GGAGGCCCTCTGACCTGTGAGAGGACGCACTTACTAGCTCTATGGATAGTGAAGTGG 1587
Db 423 GGGGGCCCTTAACTCTGTGAGAAGGATGAACTTACTACCTCTACGGGATTGTAAGCTGG 482
Qy 1588 GGCCTGGATGTGGGAAGGAGCCAGGCTCTACACCAAGTTACCAATTTCTGAAATGG 1647
Db 483 GGCAGGAATGTGGGAAGGAGCCAGGATCTACACTCAAGTCACCAAGTTCTGAAATGG 542
Qy 1648 ATCAAGCCACCATCAAAAGTGAAGTG 1675
Db 543 ATAAAGACCACCATGCACAGGAGGCTG 570

RESULT 13
BF788188 929 bp mRNA linear EST 12-JAN-2001
LOCUS 602113411F1 NCI CGAP_Kid14 Mus musculus cDNA clone IMAGE:4241642
DEFINITION 5', mRNA sequence.
ACCESSION BF788188
VERSION BF788188.1 GI:12093224
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 929)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-rc@mail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLM9858 row: e column: 03
High quality sequence stop: 686.
Location/Qualifiers
1..929
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:4241642"
/lab_host="DH10B (T1 phage-resistant)"
/clone_lib="NCI_CGAP_Kid14"
/note="Organ: Kidney; Vector: pCMV-SPORT6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.75 kb. Constructed by Life
Technologies. Note: this is a NCI_CGAP Library."

FEATURES
source
1..929
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:4241642"
/lab_host="DH10B (T1 phage-resistant)"
/clone_lib="NCI_CGAP_Kid14"
/note="Organ: Kidney; Vector: pCMV-SPORT6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.75 kb. Constructed by Life
Technologies. Note: this is a NCI_CGAP Library."

ORIGIN
Query Match 24.7%; Score 416.4; DB 10; Length 929;
Best Local Similarity 75.9%; Pred. No. 2.9e-107;
Matches 529; Conservative 0; Mismatches 162; Indels 6; Gaps 1;

Qy 1 ATGTTGCCAGATGCTGATCTCCATGTTCTGCTGTTAATGCTCTGTTGGGAAGACA 60
Db 44 ATATTGTGAGATGTAGGTGTTCCGTGCTCTAATCGCTGTTGGGAAGTCA 103
Qy 61 GCCTGTGGGTTCCTCCCTGATGTCCTTTATTGGAAGCCTGGAGCCAGACTGGACCCCTGAC 120
Db 104 GTCATGGGCTCTCATGTATGTCCTTCATTGCGCCCCAGAGCCAGATNGACCCCGAT 163
Qy 121 CAGTATGATTACAGCTACGAGGATTAATCAGGAAGAGAACACAGTAGTACACTTACC 180
Db 164 GACTATTACTACAGCTATGAGCAGTCCAGCCAGAGACCCAGAGCCCGAGTGTACGAGAC 223
Qy 181 CATGCTGAGATCCTGACTGGTACTACACTGAGGACCAAGCTGATCCATGCGAGCCCAAC 240
Db 224 ACCCTTGAGAACCCCGACTGTTACT-----ATGAGAGCATGATCCATGCCAGTCCAAC 277
Qy 241 CCCTGTGAACACCGTGGGACTCCCTGCTCCATGAGGAGCACCTTACATGACAGTCCCTG 300
Db 278 CCCTGTGAACACCGCGGGGACTGTATCATCAGAGGGGATACCTTCAGTTGCGAGTCCCA 337
Qy 301 GCTCTTCTCTCGGAATAAGTGTCAAGAGTCAAAATACGTGCAAGGACAAACCCATGT 360
Db 338 GCCCTCTCTCGGGGAGCCGCTGCCAGATGACAAACAAAGTGCAGAGCAACCCATGT 397
Qy 361 GGCGGGGCCAATGTCTCAATTACCCAGAGTCTCCCTAATTACCGCTGTGTCTGTAACAC 420
Db 398 GTCCATGCTGATTGCTCTCATACCCAGAGACCCCTTACTACCGCTGTGCTGCAATAAC 457
Qy 421 CTTTACAGAGTCCAGCTCTCCAGTGTCTTCTGTATGATGAGGCCAAACCCCTGCCAG 480
Db 458 CTTTACAGGAGACCAAGTCTCTCAAGTCTTCGGCATGAGGCCAAACCCCTGCCAG 517
Qy 481 AATGGGCTACCTGCTCCCGGCATAAAGCGAGATCCAAAGTTCACCTGTGCTGTCCCGAC 540
Db 518 AATGGCGAGTCTGTTCCCGACACAGACGAGATCCAGGTTTACCTGTGCTGTCCAGAC 577
Qy 541 CAGTCAAGGGGAATTCGTGAATAGTTCCTGATGATGCTGCTATGTTGGCGATGGCTAC 600
Db 578 CAGTATAAGGGGAATTCGTGAATAGTTCCTGATGATGCTGCTATGTTGGCGATGGCTAC 637
Qy 601 TCTTACCGAGGAAATGAATAGGACAGTCAACAGCATCGCTGCTCTTACTGGAATCC 660
Db 638 TCTTACCGAGCAAGTGAATAGGACAGTCCACAGAAACCATGCTTTACTGGACTCC 697
Qy 661 CACTCTCTTTCAGGAGATTAACATGTTTATGG 697
Db 698 CACTCTCTTTCAGGAGACTTATCATGTTCTCTG 734

RESULT 14
CF169114 665 bp mRNA linear EST 25-JUL-2003
LOCUS B0809C08-5 NIA Mouse Newborn Kidney cDNA Library (Long 1) Mus
DEFINITION musculus cDNA clone NIA:B0809C08 IMAGE:30468319 5', mRNA sequence.
ACCESSION CF169114
VERSION CF169114.1 GI:33278663
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 665)
AUTHORS Piao, Y., Ko, N.T., Lim, M.K. and Ko, M.S.H.
TITLE Construction of long-transcript enriched cDNA libraries from
submicrogram amounts of total RNAs by a universal PCR amplification
method
JOURNAL Genome Res. 11 (9), 1553-1558 (2001)
MEDLINE 21429098
PUBMED 11544199
COMMENT Contact: Dawood B. Dudekula
Laboratory of Genetics
National Institute on Aging/National Institutes of Health
333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6920, USA

Email: cna@lgsun.grc.nia.nih.gov
 Plate: B0809 row: C column: 08
 Seq primer: M13 Reverse
 High quality sequence stop: 665
 POLYA-No

FEATURES

Location/Qualifiers
 1. 665
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="niaEST:B0809C08-5"
 /db_xref="taxon:10090"
 /clone="NTA:B0809C08 IMAGE:30468319"
 /dev_stage="Newborn Kidney"
 /lab_host="DH10B"
 /clone_lib="NIA Mouse Newborn Kidney cDNA Library (Long 1)"
 /note="Vector: pCMV-SPORT6 (Invitrogen); Site 1: SalI; Site 2: NotI; Mouse cDNA project by the Laboratory of Genetics, National Institute on Aging (NIA), Intramural Research Program, NIH (http://lgsun.grc.nia.nih.gov/cDNA). In brief, double-stranded cDNAs were synthesized with an oligo(dT) primer (Invitrogen):
 5'-TGACTAGTTCAGATCGAGCGCGCCGCTTTT-3' from 26 ug of total RNA, treated with T4 DNA polymerase, and purified by ethanol-precipitation. The cDNAs were ligated to linc-linker L1-SalI, purified by phenol/chloroform, and separated from free linkers by Centricon 100. Then, the cDNAs were amplified by long-range high fidelity PCR using Ex Taq polymerase (Takara) with a primer SalI-S. The products were purified by phenol/chloroform and Centricon 100. The cDNAs were digested with SalI and NotI enzymes and cloned into SalI/NotI site of pCMV-SPORT6 plasmid vector. The DH10B E. coli host was transformed with the ligation mixture by the standard chemical method. The average insert size is about 3.0 Kb. The library was constructed by Yulan Piao."

ORIGIN

Query Match 23.7%; Score 399.4; DB 14; Length 665;
 Best Local Similarity 76.8%; Pred. No. 1.7e-102;
 Matches 503; Conservative 0; Mismatches 145; Indels 6; Gaps 1;
 QY 1 ATGTTTGCAGAGTCTGATCTCATCTTCTGCTTTAATGGCTGTGGTGGGAAGACA 60
 DB 17 ATATTGTGAGGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 76
 QY 61 GCCTGTGGTCTCCCTGATCTCTTATTGGAAGCTGACCCAGAGTGGACCCCTGAC 120
 DB 77 GTCAATGGGCTCTCACTGATGCTCTTCAATGGGCCCCAGACCCAGTGTCCAGCCGAT 136
 QY 121 CAGTATGATTACAGCTACGAGGATTATTAATCAGGAAGAGAACACAGTAGCAGCACTTACC 180
 DB 137 GACTATTACTACAGCTATGAGCAGTCCAGCCAGAGAGAGAGAGAGAGAGAGAGAGAG 196
 QY 181 CATGCTGAGATCTGACTGTGCTACTGAGGAGCAAGCTGATCCATGCGCCGACCC 240
 DB 197 ACCCTGTGAACCCGACTGGTACT-ATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 250
 QY 241 CCCTGTGAACAGGCTGGGAGTCCCTGCTCCATGGGAGCAGCTTCAATGAGTGGCTG 300
 DB 251 CCCTGTGAACAGGCGGAGCTGTATCATCAGAGGGGATACCTTCAGTTGAGCTGCCA 310
 QY 301 GTCCTTTCTCTGGGAATAGTGTGAGAAAGTGGAAATACGTGGAAGAGACACCCATGT 360
 DB 311 GCGCCCTTCTCGGGAGCGGTGCCAGACTGACAAACAAAGTGAAGAGACACCCATGT 370
 QY 361 GCGCGGGGCAATGTCTATTACCCAGAGTCTCTCTACTACCGTGTGTCTGTAACAC 420
 DB 371 GTCCATGTTGATGCTCTATTACCCAGAGACCCCTACTACCGTGTGTCTGTAACATAC 430
 QY 421 CTTTACAGAGTCCAGCTGCTCCCAAGTGGTTCTGTATGAGGCCCAAAACCCCTGCCAG 480

Db 431 CTTTACAGGACAGACTGCTCCAAAGTGTCTCGGATCGAGCCAAACCCCTGCCAG 490
 QY 481 AATGGGGTACCTGCTCCCGCATTAAGCGGAGATCCAAAGTTCACCTGTGCTGCTCCGAC 540
 DB 491 AATGGCGAGTCTGTTTCCCGACACAGAGAGATCCAGGTTTACCTGTGCTGCTCCGAC 550
 QY 541 CAGTTCAAGGGGAATTTCTGTGAATAGTGTCTGATGACTGTCTATGTTGGGATCGCTAC 600
 DB 551 CAGTATAAGGGGAATTTCTGTGAATAGTGTCTGAGAGTGTCTATGTTGGTATCGCTAC 610
 QY 601 TCTTACCGAGGAAATGAATAGGACAGTCAACAGCATGCGTGTGCTTACTTGA 655
 DB 611 TCTTACCGAGGCAAGTGAAGACAGTCAACAGAGACCCATGCTTACTTGA 665

RESULT 15
 BI332440

LOCUS 602980841F1 NCI_CGAP_Li9 Mus musculus cDNA clone IMAGE:5133587 5',
 DEFINITION mRNA sequence.

ACCESSION BI332440

VERSION BI332440.1 GI:15017097

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 791)
 NIH-MGC http://mgi.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 UNPUBLISHED (1999)
 CONTACT: Robert Strausberg, Ph.D.
 Email: cgabbs@mail.nih.gov

COMMENT Tissue Procurement: Jeffrey E. Green, M.D.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 DNA Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLAM11328 row: a column: 12
 High quality sequence stop: 787.

FEATURES

source
 1. 791
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="FVB/N"
 /db_xref="taxon:10090"
 /clone="IMAGE:5133587"
 /lab_host="DH10B (T1 phage-resistant)"
 /clone_lib="NCI_CGAP_Li9"
 /note="Organ: liver; Vector: pCMV-SPORT6; Site 1: NotI; Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.9 kb. Constructed by Life Technologies. Note: this is a NCI_CGAP Library."

ORIGIN

Query Match 23.6%; Score 396.4; DB 12; Length 791;
 Best Local Similarity 75.8%; Pred. No. 1.4e-101;
 Matches 520; Conservative 0; Mismatches 161; Indels 7; Gaps 2;
 QY 1 ATGTTTGCAGGATGCTGATCTCCATGTTCTGCTTTAATGGCTGTGGGAAGACA 60
 DB 23 ATATTGTGAGGATGTTGTTGTTCCGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 82
 QY 61 GCCTGTGGTTCCTCCTGATGTTTATTGGAAGCTGGACCCAGACTGGACCCCTGAC 120
 DB 83 GTCAATGGGCTCTCACTGATGCTCTTCAATGCGCCCCCAGAGCCAGATGAGACCCCGAT 142
 QY 121 CAGTATGATTACAGCTACGAGGATTATTAATCAGGAAGAGAACACAGTAGCAGCACTTACC 180
 DB 143 GACTATTACTACAGCTATGAGCAGTCCAGCCAGAGAGAGAGAGAGAGAGAGAGAGAG 202

QY	181	CATGCTGAGAACTCTGACTGGTACTACACTGAGGACCAAGCTGATCCATGCCAGCCCCAAC	240
Db	203	ACCCCTGAGAACCCCGACTGCTACT-----ATGAGACGATGATCCATGCCAGTCCAAC	256
QY	241	CCCTGTGAACACCGTGGGAGCTGCTCGTCCATGGGAGCACCTTCACATGCAGCTGCCTG	300
Db	257	CCCTGTGAACACCGCGGGGACTGTATCATCAGAGGGGATACCTTCAGTTGCAGCTGCCCA	316
QY	301	GCTCCTTTCTCTGGGAATTAAGTGTCAAGAAAGTGCAGAAATACGTGCAAGGACCAACCCATGT	360
Db	317	GCCCCCTTCTCGGGGAGCCGGTCCAGACTGCACAAAACAAGTGCAGGACAAACCCATGT	376
QY	361	GGCCGGGGCCAAATGTCTCAATTACCCAGAGTCTCTCCCTACTACCGTGTGTCTGTAAACAC	420
Db	377	GTCCATGGTGATTGCCCTCATTTACCCAGAAAGCACCCCTACTACCGTGTGTGCCCTGCAATAC	436
QY	421	CCTTACACAGGTCCTCAGCTGCTCCCAAGTGGTTCTGTATGCAGGCCCAACCCCTGCCAG	480
Db	437	CCTTACACGGGACACAGCTGCTCCAAAGTCTTCCGGCATGCAGGCCAAACCCCTGCCAGA	496
QY	481	AATGGGGCTACCTGCTCCCGGCATAAGCGGAGATCCAAAGTTCACCTGTGCTGTCCCGAC	540
Db	497	ATGCCGGAGTCTGTTCCTCCGACACAGACGAGATCCAGGTTTACCTGTGCTGTCCAGAC	556
QY	541	CAGTTCAGGGGAAATCTGTGAA-ATAGTTCTGATGACTGCTATGTTGGGATGGCTA	599
Db	557	CAGTATAAGGGGAAATCTGTGAACATAGTCCGGACGACTGTATTATGTCGGTGTGGCTA	616
QY	600	CTCTTACCGAGGAAATGAATAGGACAGTCAACCCAGCATGGTGCCTTTACTGGAACTC	659
Db	617	CTCTTACCGAGCAAAAGTGAGTAAGACAGTTCAACAGAACCCATGCCCTTTACTGGAACTC	676
QY	660	CCAGCTCCTCTTGCAGGAGAAATTACAAC	687
Db	677	CCAGCTCCTCTTGCAGGAGACTTATAAC	704

Search completed: May 25, 2004, 10:42:51
Job time : 3056 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 25, 2004, 05:49:14 ; Search time 91.5 Seconds
(without alignments)
10207.457 Million cell updates/sec

Title: US-09-912-559-1
Perfect score: 1683
Sequence: 1 atgtttgccaggatgtctga.....aaagtgaagtggtctctaa 1683

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA.*

- 1: /cgn2_6/ptodata/2/ina/5A COMB.seq.*
- 2: /cgn2_6/ptodata/2/ina/5B COMB.seq.*
- 3: /cgn2_6/ptodata/2/ina/6A COMB.seq.*
- 4: /cgn2_6/ptodata/2/ina/6B COMB.seq.*
- 5: /cgn2_6/ptodata/2/ina/PCTUS COMB.seq.*
- 6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	117.6	7.0	2033	1	US-08-148-910-14
2	117.6	7.0	2033	1	US-08-448-937A-14
3	96.2	5.7	970	1	US-08-148-910-3
4	96.2	5.7	970	1	US-08-448-937A-3
5	75.4	4.5	1065	1	US-08-427-640-1
6	75.4	4.5	1065	1	US-08-427-640-5
7	75.4	4.5	1068	1	US-08-427-640-3
8	75.4	4.5	1137	4	US-09-553-498-9
9	75.4	4.5	1137	4	US-09-618-869-9
10	75.4	4.5	1137	4	US-08-811-949-48
11	75.4	4.5	1955	2	US-08-883-795A-39
12	75.4	4.5	2457	6	5344773-1
13	75.4	4.5	2544	4	US-09-703-695A-3
14	75.4	4.5	7360	1	US-08-286-740-1
15	75.4	4.5	7360	5	PCT-US95-09576-1
16	73.8	4.4	329	1	US-08-148-910-2
17	73.8	4.4	329	1	US-08-448-937A-13
18	73.8	4.4	329	1	US-08-448-937A-2
19	73.8	4.4	329	1	US-08-448-937A-13
20	73.8	4.4	1065	2	US-08-811-949-60
21	73.8	4.4	1068	1	US-08-137-116-2
22	73.8	4.4	1068	1	US-08-427-640-7
23	73.8	4.4	1068	1	US-08-811-949-44
24	73.8	4.4	1068	2	US-08-811-949-46
25	73.8	4.4	1068	2	US-08-811-949-52
26	73.8	4.4	1068	2	US-08-811-949-58
27	73.8	4.4	1068	6	5223256-3

28	73.8	4.4	1163	2	US-08-558-269-5
29	73.8	4.4	1163	3	US-09-410-882-5
30	73.8	4.4	1170	2	US-08-811-949-64
31	73.8	4.4	1170	2	US-08-811-949-66
32	73.8	4.4	1314	2	US-08-811-949-50
33	73.8	4.4	1314	2	US-08-811-949-54
34	73.8	4.4	1314	2	US-08-811-949-56
35	73.8	4.4	1419	2	US-08-811-949-62
36	73.8	4.4	1738	6	5200340-1
37	73.8	4.4	1848	3	US-08-814-412-10
38	73.8	4.4	1974	2	US-08-811-949-38
39	73.8	4.4	2101	2	US-08-811-949-42
40	73.8	4.4	2162	1	US-08-119-512-3
41	73.8	4.4	2162	1	US-08-488-015B-3
42	73.8	4.4	2162	1	US-08-488-015B-25
43	67.8	4.0	1724	6	5200340-5
44	67.8	4.0	2497	6	5185259-2
45	66.4	3.9	2544	3	US-09-518-046-3

ALIGNMENTS

RESULT 1
US-08-148-910-14
; Sequence 14, Application US/08148910
; Patent No. 5466593
; GENERAL INFORMATION:
; APPLICANT: Takeshi SHIMOMURA et al.
; TITLE OF INVENTION: NO. 5466593el Protein and Gene Encoding Said Protein
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack
; STREET: 805 Fifteenth Street, N.W., #700
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch,
; MEDIUM TYPE: 500 Kb Storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; FILING DATE: US/08/148,910
; APPLICATION NUMBER: 5466593ember 5, 1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Warren M. Cheek, Jr.
; REGISTRATION NUMBER: 33,367
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-8850
; TELEFAX: 202-371-8856
; TELEX:
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2033 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; ORIGINAL SOURCE:
; ORGANISM: human
; IMMEDIATE SOURCE:
; LIBRARY: Pre-made Lambda phage Library,
; LIBRARY: human liver(49, male) cDNA Library (Stratagene)
US-08-148-910-14

QY 529 GCCTGTCCCGACCAAGTTCAAGGGGAAATTCGTGAATAGTTCTGATGA---CTCCTAT 585
 Db 802 GCCTGTCCCGACCAAGTTCTGCTGAGCGCTCTGCAACATCGAGCCTGATGAGCGCTGCTTC 861
 QY 586 GTTGGCGATGCTACTCTTACCGAGGGAATGAATAGGACAGTCAACAGCATGCGTC 645
 Db 862 TTGGGGAACGGCACTGGGTACCGTGGGTGGGCGGACACCTCAGCTCGGGCCCTCAGCTGC 921
 QY 646 CTTTACTGGAACTCCCACTCTCTTCAGAGAAATTTACAACATGTTTATGAGGATGCT 705
 Db 922 CTGGCTTGGAACTCCGATCTGCTTACAGAGAGTGCACGTGGACTCCGTGGCGCGCG 981
 QY 706 GAAACCATGGATTTGGGGAACACAAATTTCTGAGAAACCCAGATCGGACGAAAGCCC 765
 Db 982 GCCCTGTGGGCTTGGGCGGCGGCTGCTTCTGCGGAAATCGGACAAATGACGAGAGCCC 1041
 QY 766 TGGTGTCTTATTAAGTTTACCAATGCAAGGTGAATGGGAATGCTGATGCTCAGCC 825
 Db 1042 TGGTGTCTACGT---GGTGAAGGACGCGCTCTCTTGGGAGTACTGCGGCTGGAGGC 1098
 QY 826 TGCTCAGCCAGGACGTTGCCCTACCCAGAGAAAGCCCACTGAGCCATCAACCAAGCTT 885
 Db 1099 TGCGAATCCCT-----CACGAGATCCAACTGTCCCGGATCTCTTGGCGACCCGT 1149
 QY 886 CCGGGTTTGACTCCTGTGGAAGACTGAGATAGCAGAGGAGAAATCAAGAGAATCTAT 945
 Db 1150 CTTGAGCAGCTTCCCGGGCGGCGGCTGCGGAGAGGACCAAGAGAGAGAGCTTC 1209
 QY 946 GGAGGCTTTAAGACACGCGGGCAAGCAACCCATGGCAGGCTCCCTCCAGTCTCGCTG 1005
 Db 1210 CTGGGCGCACGTATCATCGGGGCTCTCTCTGCTGCGGCTGCGACCCCTG----- 1262
 QY 1006 CCTGTGACCATCTCCATGCCCCAGGCGCACTTCTGTGTGGGGCGCTGATCCACCCCTGC 1065
 Db 1263 --GCTGGCGGCATCTACATCGGGAACAGCTTCTGCGCGGAGCCTTGTTCCACCTGC 1320
 QY 1066 TGGGTGCTCACTGTGTCCTCACTGCACCGACATATAAAACACGACATCTAA-----AGGTG 1119
 Db 1321 TGGGTGCTGCTGCGCGCACTGTCTTCCACAGCCCGGCGGAGAGAGCTCTCGGTG 1380
 QY 1120 GTGCTAGGGGACAGGACTGAGAAAGAAATTTATGAGCAGAGTTTAGGGTGGAG 1179
 Db 1381 GTGCTGGGCGACACTTCTTCAACCGCAGCAGGAGTGCAGCAGACCTTCGGCATCGAG 1440
 QY 1180 AAGATATTCAAGTACAGCAGCTACAAATGAAGAGATGAGATTCGCCCAATGATTTGCA 1239
 Db 1441 AAGTACATCCCGTACACCTGTACTCGTGTTCACCCCGGAGCCAC---GACCTGTC 1497
 QY 1240 TTGCTCAAGTTAAAGCAGTGGTCACTGTGCTCTAGAAATCAAAATAGTGAAGACT 1299
 Db 1498 CTGATCCGGCTGAGAAGAAAGGAGCCGCTGTGCCACACGCTCGCAGTTCGTGAGGCC 1557
 QY 1300 GTGTGCTTGCCTGAT-----GGGTCTCTTCCCTCTGGAGTGAAGTGCACATCTCTGGC 1353
 Db 1558 ATCTGCTGCCGAGCCCGGAGCAGCTTCCCGCAGGACACAGTGCAGATTCGGGGC 1617
 QY 1354 TGGGG-----TGTTACGAACACAGAAAGGGTCCCGCAGCTCTCTGGATGCCAAAGTC 1407
 Db 1618 TGGGGCCACTTGGATGAGAACTGAGCGGCTACTCCAGCTCCCTCGGGAGCCCTGTGTC 1677
 QY 1408 AAGCTGATGCAACACTTTGTGCACTCCCGCAACTCTATGACCAATGATGATGAC 1467
 Db 1678 CCCCTGGTCCGCAACCAAGTGCAGCAGCCCTGAGGTTACGGCGCGGACATCAGCCCC 1737
 QY 1468 AGTATGATCTGTGAGGAAATCTTTCAGAAACCTGGGCAAGACCTGCGCAGGCTGACTCT 1527
 Db 1738 AACATGCTCTGTSCCGGTACTT---CGACTGCAAGTCCGAGCGCTTCCCGAGGGGACTCA 1794
 QY 1528 GGAGGCGCCCTGACCTGTGAGAGGAGCGCACCTACTAGTCTATGAGGATAGTGAAGTGG 1587
 Db 1795 GGGGGGCCCCCTGGCCTGCGAGAGAAACGGCGTGGCTTACTCTTACGGCATCATCAGCTGG 1854

QY 1588 GGCCTGGAGTGTGG-----AAGAGCCGAGGGTCTACACCCCAAGTTACCAAAATTCCTG 1641
 Db 1855 GGTGACGCTCGGGCGGCTCCACAGCCGGGGTCTACACCCCGTGGCCACTATGTG 1914
 QY 1642 AATTGGATCAAAAGCCACCAT 1661
 Db 1915 GACTGGATCAACGACCGGAT 1934

RESULT 3
 US-08-148-910-3
 ; Sequence 3, Application US/08148910
 ; Patent No. 546593
 ; GENERAL INFORMATION:
 ; APPLICANT: Takeishi SHIMOMURA et al.
 ; TITLE OF INVENTION: No. 546593el Protein and Gene Encoding Said Protein
 ; NUMBER OF SEQUENCES: 14
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSES: Wenderoth, Lind & Ponack
 ; STREET: 805 Fifteenth Street, N.W., #700
 ; CITY: Washington
 ; STATE: D.C.
 ; COUNTRY: U.S.A.
 ; ZIP: 20005
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette, 5.25 inch,
 ; MEDIUM TYPE: 500 Kb Storage
 ; COMPUTER: IBM Compatible
 ; OPERATING SYSTEM: MS-DOS
 ; SOFTWARE: Wordperfect
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/148,910
 ; FILING DATE: No. 546593ember 5, 1993
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER:
 ; FILING DATE:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Warren M. Cheek, Jr.
 ; REGISTRATION NUMBER: 33,367
 ; REFERENCE/DOCKET NUMBER:
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 202-371-8850
 ; TELEFAX: 202-371-8856
 ; TELEX:
 ; INFORMATION FOR SEQ ID NO: 3:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 970 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: cDNA
 ; ORIGINAL SOURCE:
 ; ORGANISM: human
 ; IMMEDIATE SOURCE:
 ; LIBRARY: Pre-made Lambda phage Library, human liver (49, male)
 ; LIBRARY: cDNA Library (Stratagene)
 ; US-08-148-910-3

Query Match 5.7%; Score 96.2; DB 1; Length 970;
 Best Local Similarity 53.0%; Pred. No. 8.8e-19;
 Matches 359; Conservative 0; Mismatches 288; Indels 30; Gaps 6;

QY 1009 CTGACCATCTCCATGCCCGGAGGCGGCTTCTGTGTGGGCGCTGATCCACCCCTGCTGG 1068
 Db 201 CTGGCCGCCATCTACATCGGGGACAGCTTCTGCGCGGAGCCTGTGTCACACCTGCTGG 260
 QY 1069 GTGCTCACTGCTGCCCTGACCGACATATAAAACACGACA-----TCTAAGGTGCTG 1122
 Db 261 GTGCTGTGCGCGGCCACTGCTTCTCCACAGCCCCCCCCAGGACAGCGTCTCCGTGGTG 320
 QY 1123 CTAGGGACACGAGGACCTGAAGAAAGAAATTTTCATGAGCAGAGCTTTAGGGTGGAGAAG 1182

Db 321 CTGGCCAGCAGCTTCTTCAACCGCAGCAGCGAGCTGACGACAGCTTCGGCATCGAGAAG 380
Qy 1183 ATATTCAATGACAGCAGCTTACATGAAGAGATGAGATTCCTCCCAATGATATTCATTTG 1242
Db 381 TACATCCGCTACACCTGTACTCGGTGTTCAACCCCA---GGACACAGCAGCTTCCTG 437
Qy 1243 CTAAGTTAAGCCAGTGGTGTCTACTGTCTTAGAATCCAAATACGTGAAGACTGTG 1302
Db 438 ATCCGGCTGAAGAAAGGGAGCGCTGTGCCACACGCTGCGAGTTCTGTCAGCCCATC 497
Qy 1303 TGCTTCCCTGATG-----GGTCTTTCCCTCTGGAGTGAAGTGCACATCTCTGCTGG 1356
Db 498 TGCTTCCCTGATG-----GGTCTTTCCCTCTGGAGTGAAGTGCACATCTCTGCTGG 557
Qy 1357 GG-----TGTTACAGAAACAGAAAGGGTCCCGCAGCTCTCTGATGCCAAAGTCAAG 1410
Db 558 GGCCACTTGGATGAGAAAGTGAAGCGCTACTCCAGCTCTCTGCGGAGGCCCTGTGTCCTC 617
Qy 1411 CTGATTGCCAAGCTTTGTGCAACTCCGCGCAACTCTATGACCAATGATGACAGT 1470
Db 618 CTGGTCCCGACACCAAGTGCAGCAGCCTGTAGGTTCTACGGCGCGACATCAGCCCAAC 677
Qy 1471 ATGATCTGTGCAGAAATCTTCAAGAACTTCAAGAACTGGGCAAGACACCTGCCAGGGTGA 1530
Db 678 ATGCTCTGTGCGGCTACTT---CGACTGCAAGTCCGACGCTGCCAGGGGACTCAGGG 734
Qy 1531 GGCCCTTACCTGTGAGAAAGGAGCGGACCTACTACTGTCTATGGGATGAGCTGGGCG 1590
Db 735 GGCCCTTACCTGTGAGAAAGGAGCGGCTTACTCTTACGGCATCATCAGCTGGGCT 794
Qy 1591 CTGAGTGTGGG-----AAGAGCGCAGGGTCTACACCCCAAGTTACCAAAATTCCTGAAT 1644
Db 795 GACGGCTGCGGGGCTCCACAGCGGGGCTCTACACCGCGTGCAGCACTATGTGGAC 854
Qy 1645 TGGATCAAGCCCAAT 1661
Db 855 TGGATCAAGCCGAT 871

RESULT 4

US-08-448-937A-3
; Sequence 3, Application US/08448937A
; Patent No. 5677164
; GENERAL INFORMATION:
; APPLICANT: Takeshi SHIMOMURA et al.
; TITLE OF INVENTION: No. 5677164el Protein and Gene Encoding Said Protein
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; STREET: 805 Fifteenth Street, N.W., #700
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch,
; MEDIUM TYPE: 500 Kb Storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/448,937A
; FILING DATE: May 24, 1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/148,910
; FILING DATE: No. 5677164ember 5, 1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Warren M. Cheek, Jr.
; REGISTRATION NUMBER: 33,367
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-8850

; TELEFAX: 202-371-8856
; TELEX:
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 970 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; ORIGINAL SOURCE:
; ORGANISM: human
; IMMEDIATE SOURCE:
; LIBRARY: Pre-made Lambda phage Library, human liver (49, male)
; LIBRARY: cDNA Library (Stratagene)
; US-08-448-937A-3

Query Match 5.7%; Score 96.2; DB 1; Length 970;
Best Local Similarity 53.0%; Pred. No. 8.8e-19;
Matches 359; Conservative 0; Mismatches 288; Indels 30; Gaps 6;
Qy 1009 CTGACCATCTCCATGCCCCAGGGCCACTTCTGTGTGGGGCGTGTATCCACCCCTGCTGG 1068
Db 201 CTGGCCGCCATCTACATCGGGGACAGCTTCTGCGCGGGAGCCTGTGTCACACCTGCTGG 260
Qy 1069 GTGCTCACTGTGCCCCACTGCACCGACATAAAAAACAGACA-----TCTAAAGGTGGT 1122
Db 261 GTGCTGTGGCGCGCCACTGCTTCTCCACAGCCCCCCCCAGGGACAGCGTCTCGTGGTG 320
Qy 1123 CTAGGGACCCAGGACCTGGAAGAAAGAAATTTCTAGAGCAGAGCTTTAGGTGGAGAAG 1182
Db 321 CTGGCCGAGCACTTCTTCAACCCGACGACGAGCTGTGACGAGACCTTTGGCATCGAGAAG 380
Qy 1183 ATATTCAAGTACAGCCACTACATGAAGAGATGAGATTTCCCAATATGATATTGCTATG 1242
Db 381 TACATCCGTTACCTCTGTACTCGGTGTTCAACCCCA---GCGACCAAGCCTCGTCTG 437
Qy 1243 CTCAAGTTAAAGCCAGTGGATGCTGCTCTAGAAATCCAAATACGTGAAGACTGTG 1302
Db 438 ATCCGGCTGAAGAAAGGGGACCGCTGTGCCACACGCTCGCAGTTCTGTGACGCCCATC 497
Qy 1357 GG-----TGTTACAGAAACAGGAAAGGTTCGCGCAGCTCTCTGGATGCAAAAGTCAAG 1410
Db 558 GGCCACTTGGATGAGAAAGCTGAGCGGCTACTTCCAGCTCTCTGCGGAGGCGCTGGTCCCC 617
Qy 1411 CTGATTGCCAAGCTTTGTGCAACTCCGCGCAACTCTATGACCAATGATGATGACAGT 1470
Db 618 CTGGTCCCGGACCAAGTGCAGCAGCGCTGTAGGTTCTACGGCGCGACATCAGCCCCAAC 677
Qy 1471 ATGATCTGTGAGAAATCTTCAAGAACTTGGGCAAGACACCTGCCAGGGTGAATCTGGA 1530
Db 678 ATGCTCTGTGCGGCTACTT---CGACTGCAAGTCCGACGCTGCCAGGGGACTCAGGG 734
Qy 1531 GGCCCTTACCTGTGAGAAAGGAGCGGACCTACTACTGTCTATGGGATGAGTGGGCG 1590
Db 735 GGCCCTTACCTGTGAGAAAGGAGCGGCTTACTCTTACCGCATCATCAGCTGGGCT 794
Qy 1591 CTGGAGTGTGGG-----AAGAGCGCAGGGGCTCTACACCCCAAGTTACCAAAATTCCTGAAT 1644
Db 795 GACGGCTGCGGGGCTCCACAGCGGGGCTCTACACCGCGTGCAGCACTATGTGGAC 854
Qy 1645 TGGATCAAGCCCAAT 1661
Db 855 TGGATCAAGCCGAT 871

RESULT 5

US-08-427-640-1
; Sequence 1, Application US/08427640
; Patent No. 5658798

GENERAL INFORMATION:
APPLICANT: Berg et al.
TITLE OF INVENTION: Compounds and Methods for Treatment of Thromboembolic Disorder
NUMBER OF SEQUENCES: 28
CORRESPONDENCE ADDRESS:
ADDRESSEE: Eli Lilly and Company
STREET: Lilly Corporate Center
CITY: Indianapolis
STATE: IN
COUNTRY: U.S.A.
ZIP: 46285
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.0 Mb storage
COMPUTER: Macintosh
OPERATING SYSTEM: Macintosh
SOFTWARE: Microsoft Word
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/427,640
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION NUMBER: 07/689,410
FILING DATE: 22 APRIL 1991
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1065 base pairs
TYPE: Nucleic acid
STRANDEDNESS: Double
TOPOLOGY: Linear
MOLECULE TYPE: DNA
US-08-427-640-1
Query Match 4.5%; Score 75.4; DB 1; Length 1065;
Best Local Similarity 50.8%; Pred. No. 1.7e-12;
Matches 342; Conservative 0; Mismatches 286; Indels 45; Gaps 5;
QY 1028 AGGCCACTTCTGTGGTGGGGCGCTGATCCACCCCTGTGGGTGCTCACTGTGCCCACT 1087
DB 392 AGCGTTCTCTGCGGGGCGATCTCATGCTCTCTGCTGGATCTCTCTGCCGCCCACT 451
QY 1088 GCACCGA-----CATATAAAACCGAGCATCTAAAGTGTGTAGGGACCGAGCACTGA 1141
DB 452 GCTTCCAGGAGAGGTTTCCGCCCCACCACTGACGTGATCTTGGCGAGAACATACCGGG 511
QY 1142 AGAAGAAGAAATTCATGACGAGGCTTTAGGTGGAGAGATATTCAGTACACCACT 1201
DB 512 TGTCTCCCTGGCGAGGAGCAAAATTTGAAGTCGAAAAATACATTTGTCATAAGGAAT 571
QY 1202 ACAATGAAGAGATGAGATTCCCCCAATGATATTGCAATTTGCTCAAGTTAAAGCCAGTGG 1261
DB 572 TCGAT-----GATGACACTTACGACATGACATTCGCTGCTGCAGCTGAATTCGGATT 625
QY 1262 ATGCTCATGTGCTCTAGAAATCCAAATACGTGAAGACTGTGCTGCTGCTGATGGGTCTCT 1321
DB 626 CGTCCCGCTGTGCCAGGAGAGCGGTGTCGCCACTGTGTGCTTCCCGCGCGGACC 685
QY 1322 TTCCCT-----CTGGGAGTGAGTGCCACATCTCTGCTGGGTGTTTACAGAA 1369
DB 686 TGCAGTCCCGGACTGACGCGAGTGAGCTCTCCGGCTACGGCAAGCATGAGGCTTGT 745
QY 1370 CAGGAAAGGTCCTCCAGCTCCTGGATGCCAAAGTCAAGCTGATTCGCAACATTTGT 1429
DB 746 CTCCTTTCTATTCCGAGCGGCTGAGGAGGCTCATGTCCAGCTGTACCCATCCAGCCGCT 805
QY 1430 GCAACTCCCGCAACTCTATGACCAATGATGATGATGATGATGATGATGATGATGATGATGAT 1489
DB 806 GCATCATCAACATTTACTTAACAGAACAGTACCCGACACATGCTGTGTGCTGGAGACA 865
QY 1490 TTCAGAAACCTGG-----GCCAGACACCTGCCAGGCTGACTCTGAGGACC 1534
DB 866 CTCGGAGCGCGGGCCCCAGGCAAACTTGACGACGCTGCCAGGCGATTCGGGAGGCC 925
QY 1535 CCCTGACCTGTGAGAGGAGCGGACCTACTACGCTCTATGGGATAGTGACCTGGGCGCTGG 1594

DB 926 CCTGTGTGTCTGAACGATGCCCGCATGCTTGTGGGCGATCATCAGCTGGGCGCTGG 985
QY 1595 AGTGTGGGAGAGG-----CCAGGGGTCTACACCCAGTTACCAAAATTCCTGAATTGGA 1648
DB 986 GCTGTGGACAGAGAGGATGTCGGGGTGTGTACCAAGGTTTACCAACTACCTAGACTGGA 1045
QY 1649 TCAAAGCCACCAT 1661
DB 1046 TTCTGTGACACAT 1056
RESULT 6
US-08-427-640-5
Sequence 5, Application US/08427640
Patent No. 5658788
GENERAL INFORMATION:
APPLICANT: Berg et al.
TITLE OF INVENTION: Compounds and Methods for Treatment of Thromboembolic Disorder
NUMBER OF SEQUENCES: 28
CORRESPONDENCE ADDRESS:
ADDRESSEE: Eli Lilly and Company
STREET: Lilly Corporate Center
CITY: Indianapolis
STATE: IN
COUNTRY: U.S.A.
ZIP: 46285
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.0 Mb storage
COMPUTER: Macintosh
OPERATING SYSTEM: Macintosh
SOFTWARE: Microsoft Word
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/427,640
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION NUMBER: 07/689,410
FILING DATE: 22 APRIL 1991
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 1065 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA
US-08-427-640-5
Query Match 4.5%; Score 75.4; DB 1; Length 1065;
Best Local Similarity 50.8%; Pred. No. 1.7e-12;
Matches 342; Conservative 0; Mismatches 286; Indels 45; Gaps 5;
QY 1028 AGGCCACTTCTGTGGTGGGGCGCTGATCCACCCCTGTGGGTGCTCACTGTGCCCACT 1087
DB 392 AGCGTTCTCTGCGGGGCGATCTCATGCTCTCTGCTGGATCTCTCTGCCGCCCACT 451
QY 1088 GCACCGA-----CATATAAAACCGAGCATCTAAAGTGTGTAGGGACCGAGCACTGA 1141
DB 452 GCTTCCAGGAGAGGTTTCCGCCCCACCACTGACGTGATCTTGGCGAGAACATACCGGG 511
QY 1142 AGAAGAAGAAATTCATGACGAGGCTTTAGGTGGAGAGATATTCAGTACACCACT 1201
DB 512 TGTCTCCCTGGCGAGGAGCAAAATTTGAAGTCGAAAAATACATTTGTCATAAGGAAT 571
QY 1202 ACAATGAAGAGATGAGATTCCCCCAATGATATTGCAATTTGCTCAAGTTAAAGCCAGTGG 1261
DB 572 TCGAT-----GATGACACTTACGACATGACATTCGCTGCTGCAGCTGAATTCGGATT 625
QY 1262 ATGCTCATGTGCTCTAGAAATCCAAATACGTGAAGACTGTGCTGCTGCTGATGGGTCTCT 1321
DB 626 CGTCCCGCTGTGCCAGGAGAGCGGTGTCGCCACTGTGTGCTTCCCGCGCGGACC 685
QY 1322 TTCCCT-----CTGGGAGTGAGTGCCACATCTCTGCTGGGTGTTTACAGAA 1369

1028	QY	AGGGCCATTCTGTGTGGGGCGTGATCCACCCCTGTCTGGTGTCTACTGCTGCCCACT	1087
461	Db	ACGGTTTCCTGTGGGGGGCATACTCATCAGCTCTCTGTGGATTCTCTCTCGCCGCCCACT	520
1088	QY	GCACCGA-----CATAAAAACAGACATCTAAAGGTGTGTCTAGGGACACGAGACCTGA	1141
521	Db	GCTTCCAGGAGAGTTTCCGCCCCACCACTGACGGTGATCTTTGGCGACAACATACCCGG	580
1142	QY	AGAAAGAGAAATTTTCATGACGAGAGCTTTAGGTTGGAGAGATATTTCAAGTACAGCCACT	1201
581	Db	TGGTCCCTGGCGAGGAGCGAGAAATTTGAAGTCGAAAAATACATTGTCTCATAGGAAT	640
1202	QY	ACAATGAAGAGATGAGATTCCCCCAACAATGATATTGTCATGTCTCAAGTTAAAGCCAGTGG	1261
641	Db	TCGAT-----GATGACACTTACGACAATGACATTGCGCTGCTGCAGCTGAAATCGGATT	694
1262	QY	ATGTGTCACGTGTCTAGAAATCCAAATACGTGAAGACTGTGTGCTTGCTGATGSGTCCCT	1321
695	Db	CGTCCCGCTGTGCCCAGGAGAGCAGCGTGTCCGCACTGTGTGCCTTCCCCCGCGGACC	754
1322	QY	TTCCCT-----CTGGAGTGAAGTGCCACATCTCTGGCTGGGGTGTTCACAAA	1369
755	Db	TGCAGCTGCCGGACTGACCGGAGTGTGAGCTCTCGGCTACGGCAAGCATGAGCGCTTGT	814
1370	QY	CAGAAAAAGGTCGCCGCAGCTCTCTGTGATGCCAAAGTCAAGCTGATTOCCAACTTTGT	1429
815	Db	CTCCTTTCTATTTCGGAGCGGCTGAAGGAGGCTCATGTCCAGACTGTACCCATCCAGCGCT	874
1430	QY	GCACCTCCCGCCCACTCTATGACCCACATGATTGATGACACAGTATGATCTGTGCAGGAATC	1489
875	Db	GCACATCACACATTTACTTAAACAGACAGTCAACGACAACATGTGTGTCTGGAGCA	934
1490	QY	TTCAAAACCTGG-----GCAAGACACTGCCAGGGTGACTCTCGAGGCC	1534
935	Db	CTCGAGCGGGCGGCCCGCCAGGCAAACTTGACAGCCCTCGCACGGGCGAATTCGGGAGGCC	994
1535	QY	CCCTGACCTGTGAGAAGAGCGGCACCTACTACGCTCTATGGGATAGTGAGCTGGGGCGCTGG	1594
995	Db	CCCTGGTGTCTGAACGATGCCCGCATGACTTTTGTGGGCATCATCAGCTGGGCGCTGG	1054
1595	QY	AGTGTGGAAGG-----CCAGGGTCTACAGCCCAAGTTACCAAAATCTCTGATTTGGA	1648
1055	Db	GCTGTGGACAAAGATGTCCCGGTGTGTACACCAAGGTTTACCAACTACCTTAGACTGGA	1114
1649	QY	TCAAAGCCCACT	1661
1115	Db	TTCGTGACAACT	1127

RESIST. 9

```

US-09-618-869-9
; Sequence 9, Application US/09618869
; Patent No. 6455279
; GENERAL INFORMATION:
; APPLICANT: Ambrosius, Dortehe
; APPLICANT: Rudolph, Rainer
; APPLICANT: Schaeffner, Joerg
; APPLICANT: Schwarz, Elisabeth
; TITLE OF INVENTION: PROCESS FOR THE PRODUCTION OF NATURALLY FOLDED AND
; TITLE OF INVENTION: SECRETED PROTEINS BY CO-SECRETION OF MOLECULAR
; TITLE OF INVENTION: CHAPERONES
; FILE REFERENCE: 20381
; CURRENT APPLICATION NUMBER: US/09/618,869
; CURRENT FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: EP99114811.5
; PRIOR FILING DATE: 1999-07-29
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 9
; LENGTH: 1137
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:

```

STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
CITY: ARLINGTON
STATE: VA
COUNTRY: USA
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
FILING DATE: 05-MAR-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: OBLON, NORMAN F.
REGISTRATION NUMBER: 24,618
REFERENCE/DOCKET NUMBER: 18-966-0
TELEPHONE: 703-413-3000
TELEFAX: 703-413-2220
INFORMATION FOR SEQ ID NO: 48:
SEQUENCE CHARACTERISTICS:
LENGTH: 1314 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1311
US-08-811-949-48

Query Match
Best Local Similarity 4.5%; Score 75.4; DB 2; Length 1314;
Matches 342; Conservative 0; Mismatches 286; Indels 45; Gaps 5;

QY 1028 AGGCCACTCTGTGGGGGGGGTATCCACCCCTGCTGGGTGCTCACTGTGCCCACT 1087
DB 638 AGCGTTCCTGTGGGGGGGGTATCTATCAGCTCTCTGCTGATCTCTGCGGCCACT 697
QY 1088 GCACCGA-----CATAAAAACAGACATCTAAAGTGTGCTAGGGACCGAGCACTGA 1141
DB 698 GCTTCCAGGAGAGGTTTCCGCCCCACCTGACGGTGTATCTTGGGAGAACATACCGGG 757
QY 1142 AGAAGAGAAATTTTCATGACGAGAGCTTTAGGGTGGAGAGATATTCAGTACAGCCACT 1201
DB 758 TGGTCCCTGGGAGAGAGAGAGAGATTTGAAGTGCAGAAATACATGTCCATAGGAAT 817
QY 1202 ACAATGAAGAGATGAGATTTCCCAATGATATGCTCAATGCTCAAGTTAAAGCCAGTGG 1261
DB 818 TCGAT-----GATGACACTTACGACAAATGACATTTGGCTGCTGCAGCTGAAATCGGAT 871
QY 1262 ATGCTCACTGTCTAGAAATCCAAATACGTGAGAGCTGTGTCTTCCCTGATGGGTCT 1321
DB 872 CGTCCCGCTGTGCCAGGAGAGAGAGCGGTGGTCCGCACTGTGTCTTCCCGCGGACC 931
QY 1322 TTCCCT-----CTGGAGTGAAGTCCCACTCTCTGTGGTGGGTGTACAGAA 1369
DB 932 TGCAGTCCCGGAGCTGGAGCGAGTGTGAGCTCTCCGGCTACGGCAAGCATGAGCCCTGT 991
QY 1370 CAGGAAGAGGTCCCGCAGCTCTGATGCCAAGTCAAGTCAAGTCAAGTCAAGTCAAGT 1429
DB 992 CTCCTTTCTATTCGAGCGGCTGAAGGAGGCTCATGTCAGACTGTACCCATCCAGCCGCT 1051
QY 1430 GCACTCCCGCAACTCTATGACCAACATGATGATGATGATGATGATGATGATGATGATGAT 1489
DB 1052 GCACATCAACATTTACTTAACAGACAGTACCGACCAACATGCTGTGTGCTGAGACA 1111
QY 1490 TTCAGAAACCTGG-----GCAGACACCTGCCAGGTGACTCTGAGGCC 1534
DB 1112 CTCGAGCGCGGGGGCCCCAGGCAAACTTGCAGAGCGCTGCCAGGCGATTCGGGAGGCC 1171

1535 CCTGACCTGTGAGAAAGGACGGACCTACTACTAGTCTATGGGATAGTGAAGTGGGCGCTGG 1594
DB 1172 CCTGGTGTCTGAAACGATGGCCGCATCATTGGTGGGCATCATCAGCTGGGCGCTGG 1231
QY 1595 AGTGTGGAGAGG-----CCAGGGGTCTACACCCCAAGTTACCAAAATTCCTGAATTGA 1648
DB 1232 GCTGTGACAGAGGATGTCCTGGGTGTGTACACAAAGGTTACCAACTACCTAGACTGGA 1291
QY 1649 TCAAAGCCACCAT 1661
DB 1292 TTGCTGACACAT 1304

RESULT 11
US-08-883-795A-39
Sequence 39, Application US/08883795A
Patent No. 5985607
GENERAL INFORMATION:
APPLICANT: Delcove, Genevieve
APPLICANT: Awang, Gregor
TITLE OF INVENTION: Recombinant DNA Molecules and Expression
TITLE OF INVENTION: Vectors for Tissue Plasminogen Activator
NUMBER OF SEQUENCES: 39
CORRESPONDENCE ADDRESS:
ADDRESSEE: BERESKIN & PARR
STREET: 40 King Street West
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5H 3Y2
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
FILING DATE: 27-JUN-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Gravelle, Micheline
REGISTRATION NUMBER: 40,261
REFERENCE/DOCKET NUMBER: 7841-062
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 364-7311
TELEFAX: (416) 361-1398
INFORMATION FOR SEQ ID NO: 39:
SEQUENCE CHARACTERISTICS:
LENGTH: 1955 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
US-08-883-795A-39

Query Match
Best Local Similarity 4.5%; Score 75.4; DB 2; Length 1955;
Matches 342; Conservative 0; Mismatches 286; Indels 45; Gaps 5;

QY 1028 AGGGCCACTCTGTGGTGGGGGGTATCCACCCCTGCTGGGTGCTCACTGTGCCCACT 1087
DB 1103 AGCGTTCCTGTGGGGGGGCATCTCATCAGCTCTCTGCTGGATTTCTCTGCGGCCACT 1162
QY 1088 GCACCGA-----CATAAAAACAGACATCTAAAGTGTGCTAGGGACCGAGCACTGA 1141
DB 1163 GTTTCAGAGAGGTTTCGCCCCACCGCTGACCGTATCTTGGGAGACATACCGGG 1222
QY 1142 AGAAGAGAAATTTTCATGACGAGAGCTTTAGGGTGGAGAGATATTCAGGTACAGCCACT 1201
DB 1223 TGTCTCCTGGCGAGGAGGAGAGAAATTTGAAGTGCAGAAATATCATTTGTCATAGGAAT 1282

ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: patin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/09576
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/286740
FILING DATE: 05-AUG-1994
ATTORNEY/AGENT INFORMATION:
NAME: Lee, Wendy M.
REGISTRATION NUMBER: 00,000
REFERENCE/DOCKET NUMBER: 798PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-1994
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 7360 bases
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
PCT-US95-09576-1

Query Match 4.5%; Score 75.4; DB 5; Length 7360;
Best Local Similarity 50.8%; Pred. No. 5e-12;
Matches 342; Conservative 0; Mismatches 286; Indels 45; Gaps 5;

QY 1028 AGGGCCACTTCTGTGGTGGGGCGCTGATCCACCCCTGTGGTGTCTCACTGCTGCCCACT 1087
DB |||||
DB 2634 AGCGGTCTCTGTGGGGGCGATCACTCAGCTCTCTGTGGATTCTCTCGCGCCCACT 2693
QY 1088 GCACCGA-----CATAAAAACAGACATCTAAAGTGGTGTAGGGACCCAGGACCTGA 1141
DB |||||
DB 2694 GCTTCCAGGAGAGGTTCCGCCCCACCACCTGACGGTGATCTTGGCGAGAACATACCGGG 2753
QY 1142 AGAAGAGAAATTTCATGAGCAGAGCTTTAGGGTGGAGAGATATTCAAGTACAGCCACT 1201
DB |||||
DB 2754 TGGTCCCTGGCGAGGAGGAGCAGAAATTTGAAAGTCGAAAAATACATTTGCCATAGGAAT 2813
QY 1202 ACAATGAAAGAGATGAGATTCCCCACATGATTTGCTCAAGTTAAAGCCAGTGG 1261
DB |||||
DB 2814 TCGAT-----GATGACCTTACGACATGACATTTGCCGCTGCTCAGCTGAAATCGGATT 2867
QY 1262 ATGGTCACGTGCTCTAGAAATCCAAATACGTGAAGACTGTGTGCTTGCTGTGATGGTCTCT 1321
DB |||||
DB 2868 CGTCCCGCTGTGCCAGGAGAGCAGCGTGGTCCGCACTGTGTGCTTCCCGCGCGGACC 2927
QY 1322 TTCCCT-----CTGGGAGTGAAGTCCACATCTCTGGCTGGGGTGTACAGAA 1369
DB |||||
DB 2928 TGAGCTGCCGAGCTGGAGAGTGTGAGCTCTCGGCTACGGCAAGCATGAGGCTTGT 2987
QY 1370 CAGGAAAGGGTCCCGCCAGCTCTGGATGCCAAAGTCAAGCTGATGCCAACACTTTGT 1429
DB |||||
DB 2988 CTCCTTTCTATTCCGAGCGGCTGAAGAGGCTCATGTGACTGTACCCATCCAGCCGCT 3047
QY 1430 GCAACTCCCGCACTCTATGACCAATGATTTGATGACAGATGATCTGTGAGGAATC 1489
DB |||||
DB 3048 GCACATCAAAATTTACTTAACAGAACAGTACCAGCAACATGCTGTGTGTGGAGACA 3107
QY 1490 TTCAGAAACCTGG-----GCAAGACACTTCCAGGGTGACTCTGGAGGCC 1534
DB |||||
DB 3108 CTCGGAGCGCGGGCCCCAGGCAAACTTGCACGACGCTGCCAGGGGATTCGGAGGCC 3167
QY 1535 CCTGACCTGTGAGAGGAGCGCACCTACTAGTCTATCGGATAGTACGTGGGCTGG 1594
DB |||||
DB 3168 CCTGTGTGTCTGAACGATGGCCGCATGACTTTGGTGGGCATCATCAGCTGGGSCCTGG 3227

QY 1595 AGTGTGGGAAGAGG-----CCAGGGGTCTACACCAAGTTACCAAAATTCCTGAATTGGA 1648
DB |||||
DB 3228 GCTGTGGACAGAAGGATGTCCCGGGTGTGTACACCAAGTTACCAACTACTAGACTGGA 3287
QY 1649 TCAAGCCCAACCAT 1661
DB |||||
DB 3288 TTCGTGACAACAT 3300

Search completed: May 25, 2004, 10:46:15
Job time : 98.5 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 25, 2004, 09:00:39 ; Search time 528.5 Seconds
(without alignments)
14472.614 Million cell updates/sec

Title: US-09-912-559-1
Perfect score: 1683
Sequence: 1 atgtttccaggatgtctga.....aaagtgaagtggtctctaa 1683

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 2953838 seqs, 2272363821 residues

Total number of hits satisfying chosen parameters: 5907676

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications NA:
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:
13: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:
14: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:
15: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:
16: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:
17: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:
18: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:
19: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1683	100.0	1683	9	US-09-912-559-1
2	1683	100.0	1683	16	US-10-391-215-1
3	1683	100.0	3088	9	US-09-880-107-1668
4	1681.4	99.9	1683	16	US-10-391-215-2
5	1681.4	99.9	1683	16	US-10-391-215-3
6	1679.8	99.8	1683	9	US-09-912-559-2
7	1679.8	99.8	1683	16	US-10-391-215-4
8	368.4	21.9	428	10	US-09-918-995-8242
9	262.4	15.6	264	13	US-10-425-000-17
10	120.2	7.1	451	9	US-09-864-761-11164
11	117.6	7.0	2036	9	US-09-854-456-552
12	117.6	7.0	2036	9	US-09-860-107-1612
13	117	7.0	117	9	US-09-864-761-27791
14	93.2	5.5	1302	13	US-10-087-192-590

15	93.2	5.5	2299	13	US-10-665-216-10
16	75.4	4.5	1689	9	US-09-969-271-6
17	75.4	4.5	1726	13	US-10-411-037-25
18	75.4	4.5	1726	13	US-10-411-026-25
19	75.4	4.5	1726	17	US-10-410-962-25
20	75.4	4.5	1726	17	US-10-411-049-25
21	75.4	4.5	2461	13	US-10-342-887-541
22	75.4	4.5	2461	13	US-10-172-118-541
23	75.4	4.5	2509	15	US-10-193-656-7
24	75.4	4.5	2519	9	US-09-969-271-5
25	75.4	4.5	2519	13	US-10-342-887-540
26	75.4	4.5	2519	13	US-10-172-118-540
27	75.4	4.5	2544	15	US-10-443-701-3
28	75.4	4.5	2641	9	US-09-974-298-144
29	75.4	4.5	2859	10	US-09-814-353-19992
30	74.4	4.4	3186778	13	US-10-027-632-174961
31	74.4	4.4	3186778	16	US-10-027-632-174961
32	73.8	4.4	1065	10	US-09-987-455-4
33	73.8	4.4	1065	10	US-09-987-455-7
34	73.8	4.4	1128	10	US-09-987-455-2
35	73.8	4.4	1128	10	US-09-987-455-5
36	73.8	4.4	1996	16	US-10-401-077-2
37	72.8	4.3	614	9	US-09-879-792-33
38	69.2	4.1	7596	15	US-10-004-113-57
39	69.2	4.1	8064	15	US-10-004-113-56
40	68.6	4.1	8221	12	US-10-152-319A-2167
41	68	4.0	1078	10	US-09-898-837A-16
42	68	4.0	1146	13	US-10-428-275-133
43	68	4.0	1314	10	US-09-898-837A-14
44	68	4.0	1314	10	US-09-898-837A-17
45	68	4.0	1314	13	US-10-428-275-131

ALIGNMENTS

RESULT 1

US-09-912-559-1
; Sequence 1, Application US/09912559
; Patent NO. US20020142316A1
; GENERAL INFORMATION:
; APPLICANT: ROEMISCH, JUERGEN
; APPLICANT: STOEHR, HANS-ARNOLD
; APPLICANT: FRUSSLER, ANNETTE
; APPLICANT: LANG, WIEGAND
; APPLICANT: WEIMER, THOMAS
; APPLICANT: BECKER, MARGRET
; APPLICANT: NERLICH, CLAUDIA
; APPLICANT: MUTH-NAUMANN, GUDRUN
; TITLE OF INVENTION: MUTANTS OF THE FACTOR VII-ACTIVATING PROTEASE AND
; TITLE OF INVENTION: DETECTION METHODS USING SPECIFIC ANTIBODIES
; FILE REFERENCE: 06478.1457
; CURRENT APPLICATION NUMBER: US/09/912,559
; CURRENT FILING DATE: 2001-07-26
; PRIOR APPLICATION NUMBER: DE 100 36 641.4
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: DE 100 50 040.4
; PRIOR FILING DATE: 2000-10-10
; PRIOR APPLICATION NUMBER: DE 100 52 319.6
; PRIOR FILING DATE: 2000-10-21
; PRIOR APPLICATION NUMBER: DE 101 18 706.8
; PRIOR FILING DATE: 2001-04-12
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 1683
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-912-559-1

Query Match 100.0%; Score 1683; DB 9; Length 1683;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1683; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1081 GCCCACTGACCCGACATATAAACCCAGACATCTAAAGGTGGTCTAGGGGACCCAGGACCTG 1140
Db 1081 GCCCACTGACCCGACATATAAACCCAGACATCTAAAGGTGGTCTAGGGGACCCAGGACCTG 1140
QY 1141 AAGAAAGAAAGAAATTTTCATGAGCAGAGCTTTTAAAGGTGGAGAGATATTCAGGTACAGCCAC 1200
Db 1141 AAGAAAGAAAGAAATTTTCATGAGCAGAGCTTTTAAAGGTGGAGAGATATTCAGGTACAGCCAC 1200
QY 1201 TACAAATGAAGAGATGAGATTTCCCAATGATTTGCAATGCTCAAGTTAAAGCCAGTG 1260
Db 1201 TACAAATGAAGAGATGAGATTTCCCAATGATTTGCAATGCTCAAGTTAAAGCCAGTG 1260
QY 1261 GATGGTCACCTGTGCTCTAGAAATCCAAATAGCTGAAGACTGTGTGCTGTGCTGTGCTGTGCTTC 1320
Db 1261 GATGGTCACCTGTGCTCTAGAAATCCAAATAGCTGAAGACTGTGTGCTGTGCTGTGCTGTGCTTC 1320
QY 1321 TTTCCCTCTGGAGTGTGCTGACATCTCTGGCTGGGTGTACAGAAACAGGAAAGGG 1380
Db 1321 TTTCCCTCTGGAGTGTGCTGACATCTCTGGCTGGGTGTACAGAAACAGGAAAGGG 1380
QY 1381 TCCCGCAGCTCTCTGGATGCCAAAGTCAAGCTGATTCGCAACACTTTGTGCAACTCCCGC 1440
Db 1381 TCCCGCAGCTCTCTGGATGCCAAAGTCAAGCTGATTCGCAACACTTTGTGCAACTCCCGC 1440
QY 1441 CAATCTATGACCAATGATTTGATGACAGTATGATCTGTGCGAGAAATCTTCAGAAACCT 1500
Db 1441 CAATCTATGACCAATGATTTGATGACAGTATGATCTGTGCGAGAAATCTTCAGAAACCT 1500
QY 1501 GGGCAAGACACTCCAGGGTGTGCTGGAGGCCCCCTGACCTGTGAGAGAGAGCGGACCC 1560
Db 1501 GGGCAAGACACTCCAGGGTGTGCTGGAGGCCCCCTGACCTGTGAGAGAGAGCGGACCC 1560
QY 1561 TACTACGTCTATGGATAGTAGTGAGCTGGGGCTGGAGTGTGGAGAGAGCGGAGCTTAC 1620
Db 1561 TACTACGTCTATGGATAGTAGTGAGCTGGGGCTGGAGTGTGGAGAGAGCGGAGCTTAC 1620
QY 1621 ACCCAAGTTACCAATCTCTGAATTTGGATCAAGCCACCATCAAAAGTGAAGTGGCTTC 1680
Db 1621 ACCCAAGTTACCAATCTCTGAATTTGGATCAAGCCACCATCAAAAGTGAAGTGGCTTC 1680
QY 1681 TAA 1683
Db 1681 TAA 1683

RESULT 2

US-10-391-215-1
; Sequence 1, Application US/10391215
; Publication No. US20040009543A1
; GENERAL INFORMATION:
; APPLICANT: KIECHL, STEFAN
; APPLICANT: WILLEIT, JOHANN
; APPLICANT: WIEDERMAN, CHRISTIAN JOSEF
; APPLICANT: ROEMISCH, JUERGEN
; APPLICANT: WEIMER, THOMAS
; APPLICANT: FEUSSNER, ANNETTE
; APPLICANT: STOEHR, HANS-ARNOLD
; APPLICANT: DOERSAM, VOLKER
; TITLE OF INVENTION: MARBURG I MUTANT OF FACTOR VII ACTIVATING PROTEASE
; FILE REFERENCE: 06478.1457-01
; CURRENT APPLICATION NUMBER: US/10/391,215
; CURRENT FILING DATE: 2003-03-19
; PRIOR APPLICATION NUMBER: 09/912,559
; PRIOR FILING DATE: 2001-07-26
; PRIOR APPLICATION NUMBER: DE 100 36 641.4
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: DE 100 50 040.4
; PRIOR FILING DATE: 2000-10-10
; PRIOR APPLICATION NUMBER: DE 100 52 319.6
; PRIOR FILING DATE: 2000-10-21
; PRIOR APPLICATION NUMBER: DE 101 18 706.8
; PRIOR FILING DATE: 2001-04-12

QY 1 ATGTTTCCAGGATGCTGATCTCCATGTTCTGCTGTTAATGGCTCTGTTGGGAAAGACA 60
Db 1 ATGTTTCCAGGATGCTGATCTCCATGTTCTGCTGTTAATGGCTCTGTTGGGAAAGACA 60
QY 61 GCTGTGGGTTCCTCCTGATGCTTTTATTTGAAAGCCTGGACCCAGACTGACCCCTGAC 120
Db 61 GCTGTGGGTTCCTCCTGATGCTTTTATTTGAAAGCCTGGACCCAGACTGACCCCTGAC 120
QY 121 CAGTATGATTACAGCTTACAGGATTAATAATCAGGAAGAGAAACACAGTAGACACTTACC 180
Db 121 CAGTATGATTACAGCTTACAGGATTAATAATCAGGAAGAGAAACACAGTAGACACTTACC 180
QY 181 CATGCTGAGAAATCCTGACTGCTGACTACCTGAGGACCAAGCTGATCCATGCGGCCAAC 240
Db 181 CATGCTGAGAAATCCTGACTGCTGACTACCTGAGGACCAAGCTGATCCATGCGGCCAAC 240
QY 241 CCGCTGTGAACACCGTGGGACTGCTGCTGCTGAGGACACTTACATGCAAGCTGCTG 300
Db 241 CCGCTGTGAACACCGTGGGACTGCTGCTGCTGAGGACACTTACATGCAAGCTGCTG 300
QY 301 GCTCCTTTCTCTGGGAATAGTGTGCAAGAGTGCAGAAATACGTCGAGGACCAACCATGT 360
Db 301 GCTCCTTTCTCTGGGAATAGTGTGCAAGAGTGCAGAAATACGTCGAGGACCAACCATGT 360
QY 361 GCGGGGGCAATGCTCATTACCCAGAGTCCCTCTACTACCGCTGTGCTGTAAACAC 420
Db 361 GCGGGGGCAATGCTCATTACCCAGAGTCCCTCTACTACCGCTGTGCTGTAAACAC 420
QY 421 CCTTACACAGGTCCAGCTCTCCCAAGTGGTTCCTGTATGACAGGCAAAACCCCTGCCAG 480
Db 421 CCTTACACAGGTCCAGCTCTCCCAAGTGGTTCCTGTATGACAGGCAAAACCCCTGCCAG 480
QY 481 AATGGGCTTACCTGCTCCCGCATAGCGAGATCCAGTTCACCTGTGCTGCTCCGAC 540
Db 481 AATGGGCTTACCTGCTCCCGCATAGCGAGATCCAGTTCACCTGTGCTGCTCCGAC 540
QY 541 CAGTTCAAGGGGAAATCTGTGAAATAGTTCCTGATGACTGCTATGTTGCGATGCTAC 600
Db 541 CAGTTCAAGGGGAAATCTGTGAAATAGTTCCTGATGACTGCTATGTTGCGATGCTAC 600
QY 601 TCTTACCGAGGAAATGATAGACAGTCAACAGCATGCTGCTTTACTGGAATCC 660
Db 601 TCTTACCGAGGAAATGATAGACAGTCAACAGCATGCTGCTTTACTGGAATCC 660
QY 661 CACCTCTCTTGCAGGAAATCAACATGTTTATGAGGATGCTGAAACCCATGGGATT 720
Db 661 CACCTCTCTTGCAGGAAATCAACATGTTTATGAGGATGCTGAAACCCATGGGATT 720
QY 721 GGGGAACAAATTTCTGAGAAACCCAGATGCGGACGAAAGCCCTGCTTTATTTAA 780
Db 721 GGGGAACAAATTTCTGAGAAACCCAGATGCGGACGAAAGCCCTGCTTTATTTAA 780
QY 781 GTTACCAATGACAGGTGAAATGGGAATCTGTGATGCTCAGGCTGCTCAGCCAGGAC 840
Db 781 GTTACCAATGACAGGTGAAATGGGAATCTGTGATGCTCAGGCTGCTCAGCCAGGAC 840
QY 841 GTTGCTTACCCAGAGAAAGCCCACTGAGCAATCAACCAAGCTTCCGGGTTGACTCC 900
Db 841 GTTGCTTACCCAGAGAAAGCCCACTGAGCAATCAACCAAGCTTCCGGGTTGACTCC 900
QY 901 TGTGAAAGACTGATAGCAGAGAGAGATCAAGAAATCTATGAGGCTTTTAAAGAC 960
Db 901 TGTGAAAGACTGATAGCAGAGAGAGATCAAGAAATCTATGAGGCTTTTAAAGAC 960
QY 961 ACGGGGGCAAGCAACCCATGCGAGGCTCCCTCCAGTCTCTGCTGCTCAACCATCTCC 1020
Db 961 ACGGGGGCAAGCAACCCATGCGAGGCTCCCTCCAGTCTCTGCTGCTCAACCATCTCC 1020
QY 1021 ATGCCCGAGGCACTTCTGTGGTGGGCTGATCCACCCCTGCTGGGTCTCACTGCT 1080
Db 1021 ATGCCCGAGGCACTTCTGTGGTGGGCTGATCCACCCCTGCTGGGTCTCACTGCT 1080

; NUMBER OF SEQ ID NOS: 8									
; SOFTWARE: PatentIn Ver. 2.1									
; SEQ ID NO 1									
; LENGTH: 1683									
; TYPE: DNA									
; ORGANISM: Homo sapiens									
US-10-391-215-1									
Query Match 100.0%; Score 1683; DB 16; Length 1683;									
Best Local Similarity 100.0%; Pred. No. 0;									
Matches 1683; Conservative 0; Mismatches 0; Indels 0; Gaps 0;									
Qy	1	ATGTTTCCAGAGTCTGATCTCATGTTCTGTTGTTAATGGCTCTGTTGGGAAGACA	60						
Db	1	ATGTTTCCAGAGTCTGATCTCATGTTCTGTTGTTAATGGCTCTGTTGGGAAGACA	60						
Qy	61	GCCTGTGGTCTCCCTGATGCTCTTTATTGAAAGCCCTGGACCCAGACTGGACCCCTGAC	120						
Db	61	GCCTGTGGTCTCCCTGATGCTCTTTATTGAAAGCCCTGGACCCAGACTGGACCCCTGAC	120						
Qy	121	CAGTATGATTACAGTACAGAGTATTAATCAGGAAGAGACACAGTAGCAGACTTTACC	180						
Db	121	CAGTATGATTACAGTACAGAGTATTAATCAGGAAGAGACACAGTAGCAGACTTTACC	180						
Qy	181	CATGCTGAGAAATCCTGACTGTGTTACTACACTGAGGACCAAGCTGATCCATGCCGCCCAAC	240						
Db	181	CATGCTGAGAAATCCTGACTGTGTTACTACACTGAGGACCAAGCTGATCCATGCCGCCCAAC	240						
Qy	241	CCCTGTGAACACGGTGGGAGTGCCTCTGTCATGGAGCACCTTCACATGAGCTGCGTG	300						
Db	241	CCCTGTGAACACGGTGGGAGTGCCTCTGTCATGGAGCACCTTCACATGAGCTGCGTG	300						
Qy	301	GCTCCTTTCTCTGGGAATTAAGTGTGAGAAAGTGCAAAATAGCTGCAAGGACCAACCCATGT	360						
Db	301	GCTCCTTTCTCTGGGAATTAAGTGTGAGAAAGTGCAAAATAGCTGCAAGGACCAACCCATGT	360						
Qy	361	GGCCGGGCAATGTCTCATCCAGAGTCTCCCTACTACTACCGCTGTGTCTGTTAAACAC	420						
Db	361	GGCCGGGCAATGTCTCATCCAGAGTCTCCCTACTACTACCGCTGTGTCTGTTAAACAC	420						
Qy	421	CCTTACACAGTCCAGCTGCTCCCAAGTGTCTGTATGACGCAAAACCCCTGCCAG	480						
Db	421	CCTTACACAGTCCAGCTGCTCCCAAGTGTCTGTATGACGCAAAACCCCTGCCAG	480						
Qy	481	AATGGGGTACCTGCTCCCGCATAAAGCGGAGATCCAAAGTTCACTGTGCTGTCGCCGAC	540						
Db	481	AATGGGGTACCTGCTCCCGCATAAAGCGGAGATCCAAAGTTCACTGTGCTGTCGCCGAC	540						
Qy	541	CAGTTCAAGGGAAATCTGTGAATAGTCTGTGATGACTGCTATGTTGGCGATGGCTAC	600						
Db	541	CAGTTCAAGGGAAATCTGTGAATAGTCTGTGATGACTGCTATGTTGGCGATGGCTAC	600						
Qy	601	TCTTACCGAGGAAATGAATAGGACAGTCAACAGCATGCGTGCCTTTACTGGAACTCC	660						
Db	601	TCTTACCGAGGAAATGAATAGGACAGTCAACAGCATGCGTGCCTTTACTGGAACTCC	660						
Qy	661	CACCTCCTCTTGCAGGAAATTAACAATGTTTATGGAGGATGCTGAACCCATGGGATT	720						
Db	661	CACCTCCTCTTGCAGGAAATTAACAATGTTTATGGAGGATGCTGAACCCATGGGATT	720						
Qy	721	GGGGAACAATTTCTGCAGAAACCCAGATGCGACGAAAGCCCTGTGCTTTATTAAA	780						
Db	721	GGGGAACAATTTCTGCAGAAACCCAGATGCGACGAAAGCCCTGTGCTTTATTAAA	780						
Qy	781	GTTACCAATGACAGGTGAATGGGAATFACTGTGATGTTCTCAGCTGCTCAGCCAGGAC	840						
Db	781	GTTACCAATGACAGGTGAATGGGAATFACTGTGATGTTCTCAGCTGCTCAGCCAGGAC	840						
Qy	841	GTTGCTACCCAGAGGAAAGCCCTCAGCCATCAACCAAGCTTCCGGGTTTGACTCC	900						
Db	841	GTTGCTACCCAGAGGAAAGCCCTCAGCCATCAACCAAGCTTCCGGGTTTGACTCC	900						
Qy	901	TGTGAAAGACTGAGATACAGAGAGGAGATCAAGAGAATCTATGGAGGCTTTAAGAGC	960						

RESULT 3
US-09-880-107-1668
; Sequence 1668, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherif, Uwe
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 44921-5028-WO
; CURRENT APPLICATION NUMBER: US/09/880,107
; CURRENT FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; PRIOR FILING DATE: 2000-10-02

; NUMBER OF SEQ ID NOS: 3950									
; SOFTWARE: PatentIn Ver. 2.1									
; SEQ ID NO 1668									
; LENGTH: 3008									
; TYPE: DNA									
; ORGANISM: Homo sapiens									
; FEATURE:									
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 D49742									
US-09-880-107-1668									
Query Match 100.0%; Score 1683; DB 9; Length 3008;									
Best Local Similarity 100.0%; Pred. No. 0;									
Matches 1683; Conservative 0; Mismatches 0; Indels 0; Gaps 0;									
QY	1	ATGTTTGCCAGGATGCTCATCTCCATGTTCTGCTGTTAATGGCTCTGTTGGGAAAGACA	60						
DB	97	ATGTTTGCCAGGATGCTCATCTCCATGTTCTGCTGTTAATGGCTCTGTTGGGAAAGACA	156						
QY	61	GCTGTGGGTTCCTCCCTGATGCTTTTATTTGGAAGCTGGACCCAGACTGGACCCCTGAC	120						
DB	157	GCTGTGGGTTCCTCCCTGATGCTTTTATTTGGAAGCTGGACCCAGACTGGACCCCTGAC	216						
QY	121	CAGTATGATTACAGCTACGAGGATTATATCAGGAAGAGAAACACAGTAGACACTTACC	180						
DB	217	CAGTATGATTACAGCTACGAGGATTATATCAGGAAGAGAAACACAGTAGACACTTACC	276						
QY	181	CATGCTGAGATCTGACTGCTACTACACTGAGGACCAAGCTGATCCATGCCAGCCCAAC	240						
DB	277	CATGCTGAGATCTGACTGCTACTACACTGAGGACCAAGCTGATCCATGCCAGCCCAAC	336						
QY	241	CCCTGTGAACACGGTGGGACTGCCCTGCTCCATGGGAGCACCTTACATGACGCTGCCTG	300						
DB	337	CCCTGTGAACACGGTGGGACTGCCCTGCTCCATGGGAGCACCTTACATGACGCTGCCTG	396						
QY	301	GCTCCTTTCTCTGGGAATAGTGTGAGAAAGTGCAAAATACGTGCAAGGACCAACCATGT	360						
DB	397	GCTCCTTTCTCTGGGAATAGTGTGAGAAAGTGCAAAATACGTGCAAGGACCAACCATGT	456						
QY	361	GGCGGGGGCAATGCTCTATTACCCAGAGTCTCTCCCTACTACCGCTGTCTGTAAACAC	420						
DB	457	GGCGGGGGCAATGCTCTATTACCCAGAGTCTCTCCCTACTACCGCTGTCTGTAAACAC	516						
QY	421	CCTTACACAGGTCCTCCAGTGTCTCCAAAGTGGTTCCTGTATGAGGCAAAACCCCTGCCAG	480						
DB	517	CCTTACACAGGTCCTCCAGTGTCTCCAAAGTGGTTCCTGTATGAGGCAAAACCCCTGCCAG	576						
QY	481	AATGGGGCTACCTGCTCCGGCATAGCGGAGATCCAGTTCACCTGTGCTGCTCCGAC	540						
DB	577	AATGGGGCTACCTGCTCCGGCATAGCGGAGATCCAGTTCACCTGTGCTGCTCCGAC	636						
QY	541	CAGTTCAAGGGGAAATTCGTGAAATAGGTTCTGATGACTGCTATGTTGGCGATGGCTAC	600						
DB	637	CAGTTCAAGGGGAAATTCGTGAAATAGGTTCTGATGACTGCTATGTTGGCGATGGCTAC	696						
QY	601	TCTTACCGGGGAAATGATAGACAGTCAACCCAGCATGCGTGCCTTTACTGGAACTCC	660						
DB	697	TCTTACCGGGGAAATGATAGACAGTCAACCCAGCATGCGTGCCTTTACTGGAACTCC	756						
QY	661	CACCTCCTCTTGCGAGGAAATTAACAATGTTTATGAGGATGCTGAAACCCCATGGGATT	720						
DB	757	CACCTCCTCTTGCGAGGAAATTAACAATGTTTATGAGGATGCTGAAACCCCATGGGATT	816						
QY	721	GGGGAACACAAATTTCTGAGAAACCCAGATGCGGACGAAAGCCCTGGTCTTTATAA	780						
DB	817	GGGGAACACAAATTTCTGAGAAACCCAGATGCGGACGAAAGCCCTGGTCTTTATAA	876						
QY	781	GTTACCAATGACAAGGTGAAATGGGAATACGTGTGATGCTCAGCCCTGCTCAGCCAGGAC	840						
DB	877	GTTACCAATGACAAGGTGAAATGGGAATACGTGTGATGCTCAGCCCTGCTCAGCCAGGAC	936						
QY	841	GTTGCTTACCCAGAGGAAAGCCCACTGAGGCAATCAACAGTCTCCGGGGTTGACTCC	900						
DB	937	GTTGCTTACCCAGAGGAAAGCCCACTGAGGCAATCAACAGTCTCCGGGGTTGACTCC	996						

901 TGTGGAAGACTGAGATAGCAGAGAGGAAGATCAAGAGAATCTATGAGGCTTTAAGAGC 960
DB TGTGGAAGACTGAGATAGCAGAGAGAGAGATCAAGAGAATCTATGAGGCTTTAAGAGC 1056
QY ACGCGGGCAAGCAACCATGGCAGGCTCCCTCAGTCTCGCTGCTCTGACCATCTCC 1020
DB ACGCGGGCAAGCAACCATGGCAGGCTCCCTCAGTCTCGCTGCTCTGACCATCTCC 1116
QY ATGCCCCAGGCGCACTTCTGTGTGGGGCTGATCCACCCCTGCTGGGTGCTCACTGCT 1080
DB ATGCCCCAGGCGCACTTCTGTGTGGGGCTGATCCACCCCTGCTGGGTGCTCACTGCT 1176
QY GCCCAGTGCACCAATATAAACCAGACATCTTAAAGTGGTGTAGGGGACCAAGGACCTG 1140
DB GCCCAGTGCACCAATATAAACCAGACATCTTAAAGTGGTGTAGGGGACCAAGGACCTG 1236
QY AAGAAAGAGAAATTTATGAGCAGAGCTTTAGGAGTGGTGTAGGGGACCAAGGACCTG 1200
DB AAGAAAGAGAAATTTATGAGCAGAGCTTTAGGAGTGGTGTAGGGGACCAAGGACCTG 1296
QY TACAATGAAAGAGATGAGATTCCCAATATGATATTCATGCTCAAGTTAAAGCCAGTG 1260
DB TACAATGAAAGAGATGAGATTCCCAATATGATATTCATGCTCAAGTTAAAGCCAGTG 1356
QY GATGCTCACTGCTCTAGATCCAAATAGCTGAAGACTGTGCTGCTGCTGATGGTCC 1320
DB GATGCTCACTGCTCTAGATCCAAATAGCTGAAGACTGTGCTGCTGCTGATGGTCC 1416
QY TTTCCCTCTGGAGTGAAGTCCACATCTCTGGCTGGGGTGTACAGAAACAGGAAAGGG 1380
DB TTTCCCTCTGGAGTGAAGTCCACATCTCTGGCTGGGGTGTACAGAAACAGGAAAGGG 1476
QY TCCGCGCAGCTCTGAGTCCAAAGTCAAGCTGATTCGCAACACTTTGCAAACTCCCGC 1440
DB TCCGCGCAGCTCTGAGTCCAAAGTCAAGCTGATTCGCAACACTTTGCAAACTCCCGC 1536
QY CAACCTATACACACATGATTGATGACAGTATGATCTGTCAGGAAATCTTCAGAAACCT 1500
DB CAACCTATACACACATGATTGATGACAGTATGATCTGTCAGGAAATCTTCAGAAACCT 1596
QY GGGCAAGACACTGCGAGGCTGACTCTGGAGGCCCCCTGACCTGTGAGAGGACGGCACC 1560
DB GGGCAAGACACTGCGAGGCTGACTCTGGAGGCCCCCTGACCTGTGAGAGGACGGCACC 1656
QY TACTACCTCTATGGGATGAGTGGGCTGGAGTGGGAGAGGCGGAGGCTTAC 1620
DB TACTACCTCTATGGGATGAGTGGGCTGGAGTGGGAGAGGCGGAGGCTTAC 1716
QY ACCCAAGTTACCAAAATTCCTGAATTTGGATCAAAAGCCCAATCAAAAGTGAAGTGGCTTC 1680
DB ACCCAAGTTACCAAAATTCCTGAATTTGGATCAAAAGCCCAATCAAAAGTGAAGTGGCTTC 1776
QY TAA 1681
DB TAA 1779

RESULT 4
US-10-391-215-2
; Sequence 2. Application US/10391215
; Publication No. US2004009543A1
; GENERAL INFORMATION:
; APPLICANT: KIEHL, STEFAN
; APPLICANT: WILLEIT, JOHANN
; APPLICANT: WIEDERMANN, CHRISTIAN JOSEF
; APPLICANT: ROEMISCH, JUERGEN
; APPLICANT: WEIMER, THOMAS
; APPLICANT: FEUSSNER, ANNETTE
; APPLICANT: STOEHR, HANS-ARNOLD
; APPLICANT: DOERSAM, VOLKER
; TITLE OF INVENTION: MARBURG I MUTANT OF FACTOR VII ACTIVATING PROTEASE
; TITLE OF INVENTION: (FSAP) AS RISK FACTOR FOR ARTERIAL THROMBOSIS

; FILE REFERENCE: 06478.1457-01									
; CURRENT APPLICATION NUMBER: US/10/391.215									
; CURRENT FILING DATE: 2003-03-19									
; PRIOR APPLICATION NUMBER: 09/912.559									
; PRIOR FILING DATE: 2001-07-26									
; PRIOR APPLICATION NUMBER: DE 100 36 641.4									
; PRIOR FILING DATE: 2000-07-26									
; PRIOR APPLICATION NUMBER: DE 100 50 040.4									
; PRIOR FILING DATE: 2000-10-10									
; PRIOR APPLICATION NUMBER: DE 100 52 319.6									
; PRIOR FILING DATE: 2000-10-21									
; PRIOR APPLICATION NUMBER: DE 101 18 706.8									
; PRIOR FILING DATE: 2001-04-12									
; NUMBER OF SEQ ID NOS: 8									
; SOFTWARE: PatentIn Ver. 2.1									
; SEQ ID NO 2									
; LENGTH: 1683									
; TYPE: DNA									
; ORGANISM: Homo sapiens									
US-10-391-215-2									
Query Match 99.9%; Score 1681.4; DB 16; Length 1683;									
Best Local Similarity 99.9%; Pred. No. 0;									
Matches 1682; Conservative 0; Mismatches 1; Indels 0; Gaps 0;									
QY	1	ATGTTTGCAGGATCTCTGATCTCCATGTTCTCTGTTTAATGGCTCTGGTGGGAAGACA	60						
DB	1	ATGTTTGCAGGATCTCTGATCTCCATGTTCTGCTGTTAATGGCTCTGGTGGGAAGACA	60						
QY	61	GCCTGTGGTCTCCCTGATGCTTTATTGAAAGCCTGGACCCAGAGCTGACCCCTTGAC	120						
DB	61	GCCTGTGGTCTCCCTGATGCTTTATTGAAAGCCTGGACCCAGAGCTGACCCCTTGAC	120						
QY	121	CAGTATGATTACAGTACCGAGGATTAATCAGGAAGAGACACACAGTAGCACACTTACC	180						
DB	121	CAGTATGATTACAGTACCGAGGATTAATCAGGAAGAGACACACAGTAGCACACTTACC	180						
QY	181	CATGCTGGAATCCTGACTGCTACTACCTGAGGACCAAGCTGATCCATGCGAGCTGCTG	240						
DB	181	CATGCTGGAATCCTGACTGCTACTACCTGAGGACCAAGCTGATCCATGCGAGCTGCTG	240						
QY	241	CCCTGTGAACACGGTGGGAGCTGCTCTGTCATGGGAGCACCTTCACATGCGAGCTGCTG	300						
DB	241	CCCTGTGAACACGGTGGGAGCTGCTCTGTCATGGGAGCACCTTCACATGCGAGCTGCTG	300						
QY	301	GCTCCTTTCTGGAATTAAGTGTGAGAAAGTGCAAAATACGTGCAAGGACACCCATGT	360						
DB	301	GCTCCTTTCTGGAATTAAGTGTGAGAAAGTGCAAAATACGTGCAAGGACACCCATGT	360						
QY	361	GGCCGGGGCCAAATGTCTATTACCCAGAGTCTCCCTACTACCGCTGTGTCTGTAACAC	420						
DB	361	GGCCGGGGCCAAATGTCTATTACCCAGAGTCTCCCTACTACCGCTGTGTCTGTAACAC	420						
QY	421	CCTTACAGGTCCTGCTCCCGGATTAAGCGGAGATCCAAAGTTCACCTGTGCTGCCGAC	480						
DB	421	CCTTACAGGTCCTGCTCCCGGATTAAGCGGAGATCCAAAGTTCACCTGTGCTGCCGAC	480						
QY	481	AATGGGGTACTGCTCCCGGATTAAGCGGAGATCCAAAGTTCACCTGTGCTGCCGAC	540						
DB	481	AATGGGGTACTGCTCCCGGATTAAGCGGAGATCCAAAGTTCACCTGTGCTGCCGAC	540						
QY	541	CAGTTCAAGGGAAATCTGTGAATAGTTCGATGATCTGTTGGGATGGCTTAC	600						
DB	541	CAGTTCAAGGGAAATCTGTGAATAGTTCGATGATCTGTTGGGATGGCTTAC	600						
QY	601	TCTTACCGAGGAAATGAATAGGACAGTCAACAGCATGCTGCTTACTTGGAACTCC	660						
DB	601	TCTTACCGAGGAAATGAATAGGACAGTCAACAGCATGCTGCTTACTTGGAACTCC	660						
QY	661	CACCTCCTCTGCAGGAGAAATACACATGTTTATGGAGATGCTGAAACCCATGGGATT	720						
DB	661	CACCTCCTCTGCAGGAGAAATACACATGTTTATGGAGATGCTGAAACCCATGGGATT	720						

RESULT 5
US-10-391-215-3
; Sequence 3, Application US/10391215
; Publication No. US20040009543A1

QY	721	GGGGAACACAAATTTCTGCAGAAACCCAGATCGGAGCAAAAGCCCTGCTGCTTTTATAA	780
DB	721	GGGGAACACAAATTTCTGCAGAAACCCAGATCGGAGCAAAAGCCCTGCTGCTTTTATAA	780
QY	781	GTTACCAATGACAAAGTGAATGGAAATCTGTGATGTCTGAGCTGCTCAGCCCGAGAC	840
DB	781	GTTACCAATGACAAAGTGAATGGAAATCTGTGATGTCTGAGCTGCTCAGCCCGAGAC	840
QY	841	GTTGCTTACCCAGAGGAAACCCCACTGAGCCATCAACCAAGCTTCCGGGGTTTGACTCC	900
DB	841	GTTGCTTACCCAGAGGAAACCCCACTGAGCCATCAACCAAGCTTCCGGGGTTTGACTCC	900
QY	901	TGTGAAAGACTGAGATAGCAGAGAGAAAGATCAAGAGATCTATGAGAGCTTTAAGAC	960
DB	901	TGTGAAAGACTGAGATAGCAGAGAGAAAGATCAAGAGATCTATGAGAGCTTTAAGAC	960
QY	961	ACGGCGGCAAGCACCCATCGGAGGCTCCCTCCAGTCTCTCGCTGCTCTGACCATCTCC	1020
DB	961	ACGGCGGCAAGCACCCATCGGAGGCTCCCTCCAGTCTCTCGCTGCTCTGACCATCTCC	1020
QY	1021	ATGCCCCAGGGCCACTTCTGTGTGGGGCGCTGATCCACCCCTGCTGCTGGGTGCTACTGCT	1080
DB	1021	ATGCCCCAGGGCCACTTCTGTGTGGGGCGCTGATCCACCCCTGCTGCTGGGTGCTACTGCT	1080
QY	1081	GCCCACTGCACCGACATAAAACCCAGACATCTAAAGGTGCTAGGGGACCCAGGACCTG	1140
DB	1081	GCCCACTGCACCGACATAAAACCCAGACATCTAAAGGTGCTAGGGGACCCAGGACCTG	1140
QY	1141	AAGAAAGAGAAATTTTATGAGCAGAGACTTTAGGTTGGAGAGATATTCAGTACAGCCAC	1200
DB	1141	AAGAAAGAGAAATTTTATGAGCAGAGACTTTAGGTTGGAGAGATATTCAGTACAGCCAC	1200
QY	1201	TACATGAAGAGATGAGATTTCCCAATGATATTTGCAATGCTCAAGTTAAGCCAGT	1260
DB	1201	TACATGAAGAGATGAGATTTCCCAATGATATTTGCAATGCTCAAGTTAAGCCAGT	1260
QY	1261	GATGTCCTGCTCTAGAAATCCAAATACGTGAAGACTGTGTGCTTGCTGCTGATGGGTC	1320
DB	1261	GATGTCCTGCTCTAGAAATCCAAATACGTGAAGACTGTGTGCTTGCTGCTGATGGGTC	1320
QY	1321	TTTCCCTCTGGAGTGAAGTGCACATCTCTGGTGGGGTGTACAGAAACAGAAAGGG	1380
DB	1321	TTTCCCTCTGGAGTGAAGTGCACATCTCTGGTGGGGTGTACAGAAACAGAAAGGG	1380
QY	1381	TCCCGCAGCTCTGATGTCAAAGTCAAGCTGATTGCCAACACTTTGTGCAACTCCCGC	1440
DB	1381	TCCCGCAGCTCTGATGTCAAAGTCAAGCTGATTGCCAACACTTTGTGCAACTCCCGC	1440
QY	1441	CAACTCTATGACCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	1500
DB	1441	CAACTCTATGACCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	1500
QY	1501	GGGCAAGACACCTGCGAGGCTGCTGAGGCCCCCTGACCTGTCAGAGGACCGGACCC	1560
DB	1501	GGGCAAGACACCTGCGAGGCTGCTGAGGCCCCCTGACCTGTCAGAGGACCGGACCC	1560
QY	1561	TACTAGTCTATGGGATGATGAGTGGGCTGGAGTGTGGGAAGAGCCAGGGGCTTAC	1620
DB	1561	TACTAGTCTATGGGATGATGAGTGGGCTGGAGTGTGGGAAGAGCCAGGGGCTTAC	1620
QY	1621	ACCAAGTTACCAAAATCTCGAATTCGATCAAGCCACCATCAAAAGTGAAGTGGCTTC	1680
DB	1621	ACCAAGTTACCAAAATCTCGAATTCGATCAAGCCACCATCAAAAGTGAAGTGGCTTC	1680
QY	1681	TAA 1683	
DB	1681	TAA 1683	

Db 1621 ACCCAAGTTACCAATTCCTGATCGATCAAGCCACCATCAAAAGTGAAGTGCTTC 1680
 QY 1681 TAA 1683
 Db 1681 TAA 1683
 RESULT 6
 US-09-912-559-2
 ; Sequence 2, Application US/09912559
 ; Patent No. US20020142316A1
 ; GENERAL INFORMATION:
 ; APPLICANT: ROEMISCH, JUERGEN
 ; APPLICANT: STOEHR, HANS-ARNOLD
 ; APPLICANT: FEUSSNER, ANNETTE
 ; APPLICANT: LANG, WIEGAND
 ; APPLICANT: WEIMER, THOMAS
 ; APPLICANT: BECKER, MARGRET
 ; APPLICANT: NERLICH, CLAUDIA
 ; APPLICANT: MUTH-NAUMANN, GUDRUN
 ; TITLE OF INVENTION: MUTANTS OF THE FACTOR VII-ACTIVATING PROTEASE AND
 ; TITLE OF INVENTION: DETECTION METHODS USING SPECIFIC ANTIBODIES
 ; FILE REFERENCE: 06478.1457
 ; CURRENT APPLICATION NUMBER: US/09/912,559
 ; PRIORITY FILING DATE: 2001-07-26
 ; PRIOR APPLICATION NUMBER: DE 100 36 641.4
 ; PRIORITY FILING DATE: 2000-07-26
 ; PRIOR APPLICATION NUMBER: DE 100 50 040.4
 ; PRIORITY FILING DATE: 2000-10-10
 ; PRIOR APPLICATION NUMBER: DE 100 52 319.6
 ; PRIORITY FILING DATE: 2000-10-21
 ; PRIOR APPLICATION NUMBER: DE 101 18 706.8
 ; PRIORITY FILING DATE: 2001-04-12
 ; NUMBER OF SEQ ID NOS: 4
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 2
 ; LENGTH: 1683
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; US-09-912-559-2
 Query Match 99.8%; Score 1679.8; DB 9; Length 1683;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 1681; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 ATGTTGGCAGGATGCTGATCTCCATGTTCTGCTGTTAATGGCTCTGGTGGGAAGACA 60
 Db 1 ATGTTGGCAGGATGCTGATCTCCATGTTCTGCTGTTAATGGCTCTGGTGGGAAGACA 60
 QY 61 GCCTGTGGGTTCTCCCTGATGCTTTATTGGAAAGCTGGACCCAGACTGGACCCCTGAC 120
 Db 61 GCCTGTGGGTTCTCCCTGATGCTTTATTGGAAAGCTGGACCCAGACTGGACCCCTGAC 120
 QY 121 CAGTATGATTACAGCTACGAGGATTAATACAGGAAGAAACACACAGTAGCACCTTACC 180
 Db 121 CAGTATGATTACAGCTACGAGGATTAATACAGGAAGAAACACACAGTAGCACCTTACC 180
 QY 181 CATGCTGAGAACTCTGACTGCTGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 240
 Db 181 CATGCTGAGAACTCTGACTGCTGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 240
 QY 241 CCCTGTGAACACAGGTGGGACTGCCCTGCTCCATGGAGACCTTTCATGTCAGTGCCTG 300
 Db 241 CCCTGTGAACACAGGTGGGACTGCCCTGCTCCATGGAGACCTTTCATGTCAGTGCCTG 300
 QY 301 GCTCCCTTCTCTGGGATAGTGTGAAAGTCAAAATACGTGCAAGGACCAACCCATGT 360
 Db 301 GCTCCCTTCTCTGGGATAGTGTGAAAGTCAAAATACGTGCAAGGACCAACCCATGT 360
 QY 361 GGCGGGGGCCAAATGTCTCATTACCCAGAGTCCTCCCTACTACCGCTGTGTGTAAACAC 420
 Db 361 GGCGGGGGCCAAATGTCTCATTACCCAGAGTCCTCCCTACTACCGCTGTGTGTAAACAC 420

QY 421 CTTTACACAGGTCCTCCAGTGTCTCCCAAGTGGTTCCTGTATGACAGCCAAAACCCCTGCCAG 480
 Db 421 CTTTACACAGGTCCTCCAGTGTCTCCCAAGTGGTTCCTGTATGACAGCCAAAACCCCTGCCAG 480
 QY 481 AATGGGGTACCTGCTCCCGGATAGCGGATCCAGTCCACCTGTCCTGTCCTCCGAC 540
 Db 481 AATGGGGTACCTGCTCCCGGATAGCGGATCCAGTCCACCTGTCCTGTCCTCCGAC 540
 QY 541 CAGTTCAAGGGGAAATTCCTGTGAAATAGTCTCTGATGACTGCTATGTTGGCGATGCTAC 600
 Db 541 CAGTTCAAGGGGAAATTCCTGTGAAATAGTCTCTGATGACTGCTATGTTGGCGATGCTAC 600
 QY 601 TCTTACCGAGGAAATAGTATAGGACATCAACGACATGCGTTCCTTACTTGGAACTCC 660
 Db 601 TCTTACCGAGGAAATAGTATAGGACATCAACGACATGCGTTCCTTACTTGGAACTCC 660
 QY 661 CACCTCTCTTGCAGGAGAAATACAAATGTTTATGAGGATGCTGCTGAAACCCATGGATT 720
 Db 661 CACCTCTCTTGCAGGAGAAATACAAATGTTTATGAGGATGCTGCTGAAACCCATGGATT 720
 QY 721 GGGGAAACAAATTTCTGCGAAACCCAGATGCGGAGAAAGCCCTGCTTATATAA 780
 Db 721 GGGGAAACAAATTTCTGCGAAACCCAGATGCGGAGAAAGCCCTGCTTATATAA 780
 QY 781 GTTACCAATGACAAGTGAATGCGAATATCTGTGATGTCTCAGCCTGCTCAGCCAGGAC 840
 Db 781 GTTACCAATGACAAGTGAATGCGAATATCTGTGATGTCTCAGCCTGCTCAGCCAGGAC 840
 QY 841 GTTGCCCTACCCAGGAAAGCCCTGAGCGATCAACAGCTTCGGGGTTGACTCC 900
 Db 841 GTTGCCCTACCCAGGAAAGCCCTGAGCGATCAACAGCTTCGGGGTTGACTCC 900
 QY 901 TGTGGAAGACTGAGATAGCAGAGAGAGAGATCAAGAGATCTATGGAGGCTTTAAGAGC 960
 Db 901 TGTGGAAGACTGAGATAGCAGAGAGAGAGATCAAGAGATCTATGGAGGCTTTAAGAGC 960
 QY 961 ACGGCGGGCAAGCACCCATGGCAGGCGTCCCTCCAGTCTCGTCTGCTGACCATCTCC 1020
 Db 961 ACGGCGGGCAAGCACCCATGGCAGGCGTCCCTCCAGTCTCGTCTGCTGACCATCTCC 1020
 QY 1021 ATGCCCCAGGGCCACTCTCTGTGGGGCGCTCATCCACCCCTGCTGGGTGCTCACTGCT 1080
 Db 1021 ATGCCCCAGGGCCACTCTCTGTGGGGCGCTCATCCACCCCTGCTGGGTGCTCACTGCT 1080
 QY 1081 GCCCACTGCAACACATAAAGAACAGACATCTAAAGTGTGTAGGGGACAGGACCTG 1140
 Db 1081 GCCCACTGCAACACATAAAGAACAGACATCTAAAGTGTGTAGGGGACAGGACCTG 1140
 QY 1141 AAGAAAGAGAAATTTATGAGCAGAGCTTTAGGGTGGAGAGATATTCAGATACAGCCAC 1200
 Db 1141 AAGAAAGAGAAATTTATGAGCAGAGCTTTAGGGTGGAGAGATATTCAGATACAGCCAC 1200
 QY 1201 TACAATGAAGAGATGAGATTCCCAATATGATTTGCAATGCTCAAGTTAAAGCCAGTG 1260
 Db 1201 TACAATGAAGAGATGAGATTCCCAATATGATTTGCAATGCTCAAGTTAAAGCCAGTG 1260
 QY 1261 GATGGTCACTGTGCTCTAGAAATCCAAATAGTGAGAGCTGTGCTGCTGCTGGGTCC 1320
 Db 1261 GATGGTCACTGTGCTCTAGAAATCCAAATAGTGAGAGCTGTGCTGCTGCTGGGTCC 1320
 QY 1321 TTTCCCTCTGGGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1380
 Db 1321 TTTCCCTCTGGGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1380
 QY 1381 TCCCGCCAGCTCTCTGGATGCCAAAGTCAAGTGAATGCGCAACACATTTGTGCAACTCCGC 1440
 Db 1381 TCCCGCCAGCTCTCTGGATGCCAAAGTCAAGTGAATGCGCAACACATTTGTGCAACTCCGC 1440
 QY 1441 CAATCTATGACACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1500
 Db 1441 CAATCTATGACACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1500
 QY 1501 GGGCAAGACACCTGCGCAGGGTGAATCTGTGGAGGGCCCCCTGACCTGTGAGAGGACGGCAC 1560

;; PRIOR APPLICATION NUMBER: US 60/180,312
;; PRIOR FILING DATE: 2000-02-04
;; PRIOR APPLICATION NUMBER: US 60/207,456
;; PRIOR FILING DATE: 2000-05-26
;; PRIOR APPLICATION NUMBER: US 09/632,366
;; PRIOR FILING DATE: 2000-08-03
;; PRIOR APPLICATION NUMBER: GB 24263.6
;; PRIOR FILING DATE: 2000-10-04
;; PRIOR APPLICATION NUMBER: US 60/236,359
;; PRIOR FILING DATE: 2000-09-27
;; PRIOR APPLICATION NUMBER: PCT/US01/00666
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00667
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00664
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00669
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00665
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00668
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00663
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00662
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00661
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00670
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: US 60/234,687
;; PRIOR FILING DATE: 2000-09-21
;; PRIOR APPLICATION NUMBER: US 09/608,408
;; PRIOR FILING DATE: 2000-06-30
;; PRIOR APPLICATION NUMBER: US 09/774,203
;; PRIOR FILING DATE: 2001-01-29
;; NUMBER OF SEQ ID NOS: 49117
;; SOFTWARE: Annonmax Sequence Listing Engine vers. 1.1
;; SEQ ID NO 11164
;; LENGTH: 451
;; TYPE: DNA
;; ORGANISM: Homo sapiens
;; FEATURE:
;; OTHER INFORMATION: MAP TO AC006097.1
;; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.89
;; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1
;; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.8
;; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.1
;; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1
;; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.9
US-09-864-761-11164

Query Match 7.1%; Score 120.2; DB 9; Length 451;
Best Local Similarity 97.6%; Pred. No. 8.6e-29;
Matches 122; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 324 TCAGAAAGTCAAAATACGTGCAAGGACACCCCATGTGGCGGGCCCAATGTCTCATTTAC 383
Db 451 TCCTACAGTCAAAATACGTGCAAGGACACCCCATGTGGCGGGCCCAATGTCTCATTTAC 392
QY 384 CCAGAGTCTCTCTACTACCGCTGTGTCTGTAAACACCTTACACAGGTCCCAAGTGTCTC 443
Db 391 CCAGAGTCTCTCTACTACCGCTGTGTCTGTAAACACCTTACACAGGTCCCAAGTGTCTC 332
QY 444 CCAAG 448
Db 331 CCAAG 327

RESULT 11
US-09-954-456-552
; Sequence 552, Application US/09954456
; Patent No. US20020115057A1

;; GENERAL INFORMATION:
;; APPLICANT: Young, Paul
;; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using Can
;; TITLE OF INVENTION: Sets
;; FILE REFERENCE: 689290-76
;; CURRENT APPLICATION NUMBER: US/09/954,456
;; CURRENT FILING DATE: 2001-09-18
;; PRIOR APPLICATION NUMBER: US/60/233,617
;; PRIOR FILING DATE: 2000-09-18
;; PRIOR APPLICATION NUMBER: US/60/234,052
;; PRIOR FILING DATE: 2000-09-20
;; PRIOR APPLICATION NUMBER: US/60/234,923
;; PRIOR FILING DATE: 2000-09-25
;; PRIOR APPLICATION NUMBER: US/60/235,134
;; PRIOR FILING DATE: 2000-09-25
;; PRIOR APPLICATION NUMBER: US/60/235,637
;; PRIOR FILING DATE: 2000-09-26
;; PRIOR APPLICATION NUMBER: US/60/235,638
;; PRIOR FILING DATE: 2000-09-26
;; PRIOR APPLICATION NUMBER: US/60/235,711
;; PRIOR FILING DATE: 2000-09-27
;; PRIOR APPLICATION NUMBER: US/60/235,720
;; PRIOR FILING DATE: 2000-09-27
;; PRIOR APPLICATION NUMBER: US/60/235,840
;; PRIOR FILING DATE: 2000-09-27
;; PRIOR APPLICATION NUMBER: US/60/235,863
;; PRIOR FILING DATE: 2000-09-27
;; NUMBER OF SEQ ID NOS: 2276
;; SOFTWARE: PatentIn version 3.0
;; SEQ ID NO 552
;; LENGTH: 2036
;; TYPE: DNA
;; ORGANISM: Homo sapiens
US-09-954-456-552

Query Match 7.0%; Score 117.6; DB 9; Length 2036;
Best Local Similarity 49.8%; Pred. No. 1.5e-27;
Matches 607; Conservative 0; Mismatches 559; Indels 54; Gaps 10;

QY 469 AACCCCTGCCAGATGGGGCTACCTGCTCCCGGCATAGCGGAGATCCAGTTCCACCTGT 528
Db 745 AGCCCTTGCTGAACGGGGGCACCTGCCACCTGATCGTCCACCGGACCCACCGTGTGT 804
QY 529 GCCTGTCGCCAGCTTCAAGGGGAAATTCGTCAAAATAGGTTCTGTATGA---CTGCTAT 585
Db 805 GCCTGCCACCGAGCTTGGTGGAGGCTCTGCAACATCAGAGCTGTATGAGCGTGTTC 864
QY 586 GTTGGCGATGCTACTCTTACCGAGGGGAAATGAATAGGACAGTCAACAGCATGCGTGC 645
Db 865 TTGGGGAACGGCACTGGGTACCGTGGCGTGGCCAGCACCTCAGCCTCGGGCTCAGCTGC 924
QY 646 CTTTACTGGAATCCACCTCTCTTGCAGGAGAAATACAAATGTTTATGAGGATGCT 705
Db 925 CTGGCTTGAATCTCGATCTCTTACAGGAGTGCACGTGGACTCGTGGGCGCGCG 984
QY 706 GAAACCCATGGGATTTGGGGAACACAATTTCTGCAGAAACCCAGATGCGGACGAAAGCCC 765
Db 985 GCCTGCTGGGCTGGGCGCCCATGCTCTACTGCGGAATCGGACAATGACGAGAGGCC 1044
QY 766 TGGTCTTTTAAAGTTACCAATGCAAGGTGAATGGGNACTACTGTATGTCTCAGCC 825
Db 1045 TGGTCTTACGT---GGTGAAGACAGCGCTCTCTCTGGGAGTACTCGCCGCTGGAGGCC 1101
QY 826 TGCTCAGCCCGAGGAGTGTGCTTACCCAGAGGAAAGCCCCACTGAGCCATCAACCAAGTT 885
Db 1102 TGGGAATCCCT-----CACCAGAGTCCAATGTCACCGGATCTCTTGGCGACCTG 1152
QY 886 CCGGGGTTTGTACTCTCTGTGGAAGACTGAGATAGCAGAGAGAGATCAAGAGATCTAT 945
Db 1153 CTTGAGCCAGCCTTCCCGGGCGCGAGGCTTGGCGGAGGAGGCAACAGAGAGCGTTC 1212
QY 946 GGAGGCTTTAAGAGCACCGCGGCAAGCACCCATGCGAGGGGCTCCCTCCAGTCTCTCGCTG 1005

1213	CTGGGGCCAGTATCATCGGGGCTCTCTCTCGTGGCGGCTCGCACCCCTG-----	1266
Qy	CTCTGACCAATCTCCATGCCCCCAGGGCCACTCTGTGGTGGGCGGTGATCCACCCCTGC	1065
Db	--GCTGGCCGCCATCTACATCGGGGACAGCTTCTGGCCGGGAGCCTGGTCCACACCTGC	1323
Qy	TGGGTGCTCATCTGTCGCCACTCACCGACATATAAACCAGACATCTAA-----AGGTG	1119
Db	TGGGTGTTGGCGGCCCACTGCTTCTCCACAGCCCCCCAGGACAGGCTCTCCGGT	1383
Qy	GTCTAGGGACACAGGACCTGAAGAAGAAGAAATTTATGAGCAGAGCTTTAGGTGGAG	1179
Db	GTCTGGGCGAGCACTTCTTCAACCGCACAGCGAGCGTGACGACAGACCTTCGGCATCGAG	1443
Qy	AAGATATTCAAGTACAGCCACTCAATGAAGAAGATCAGATTCCTCCACATGATATGCA	1239
Db	AAGTACATCCGTAACCTGTACTCGTGTTCAACCCGACGACCAC---GACCTCGTC	1500
Qy	TTGCTCAAGTTAAAGCCAGTGGATGTCTACTGTGCTCTAGAATCCAATACGTGAAGACT	1299
Db	CTGATCCGGCTGAAGAAGAAAGGGACCGCTGTGCCACACGCTCGCAGTTCTGTGCAGCCC	1560
Qy	GTGTGCTTGCTGAT-----GGGTCTTTCCCTCTGGGAGTGAGTGCACATCTCTGGC	1353
Db	ATTGCTGTGCCAGCCCGCAGCACTTCCCGCAGGACACAAGTGCACAGATTGCGGGC	1620
Qy	TGGGG-----TGTTACAGAAACAGGAAAAGGCTCCGGCCAGCTCTCTGGATGCCAAATC	1407
Db	TGGGGCCACTTGATGAGAACGTGAGCGGCTACTCCAGTCCCTTCGGGAGGCCCTGGTC	1680
Qy	AAGCTGATTGCCAACACTTTGTGGAACTCCGGCCAACTCTATGACCAATGATGATGAC	1467
Db	CCCTGTGTCGCCACCAACAGTGCAGAGCCCTGAGTCTTACGCGCGCAGATCAGCCCC	1740
Qy	AGTATGATCTGTCAGGAAATCTTCAGAAACCTGGGCAAGACACTGCCAGGTGACTCT	1527
Db	AACATGCTCTGTGCGGCTACTT---CGACTGCAAGTCCGACGCTTCCAGGGGGACTCA	1797
Qy	GGAGGCCCCCTGACCTGTGTGAAGAGCGGCACCTTACGTTCTATGGATGATGAGCTGG	1587
Db	GGGGGGCCCCCTGGCCTTCGGAAGAAACGGCGTGGCTTACCTCTACGCGCATCATCAGCTG	1857
Qy	GGCCTGGAGTGTGG-----AAGAGGCCAGGGGTCTACACCCCAAGTTACCAATTCCTG	1641
Db	GGTGACGGCTGCGGGCGGCTTCACAGCCGGGGTCTACACCCGGTGGCCAACTATGTG	1917
Qy	AATTGGATCAAGCCACCAT	1661
Db	GACTGGATCAACACCGGAT	1937

RESULT 12

```

US-09-880-107-1612
; Sequence 1612, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherf, Uwe
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 44921-5028-WO
; CURRENT APPLICATION NUMBER: US/09/880,107
; CURRENT FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: PatentIn ver. 2.1
; SEQ ID NO 1612
; LENGTH: 2036
; TYPE: DNA

```

	;	ORGANISM:	Homo sapiens	
	;	FEATURE:		
	;	OTHER INFORMATION:	Genbank Accession No.	US20020142981A1 D14012
			US-09-880-107-1612	
		Query Match	7.0%;	Score 117.6; DB 9; Length 2036;
		Best Local Similarity	49.8%;	Pred. No. 1.Se-27;
		Matches 607; Conservative	0;	Mismatches 559; Indels 54; Gaps 10;
Qy	469	AACCCCTGCCAGAAATGGGCGTACTGTCTCCGGCATACGCGAGATCCAAGTTCACTGT 528		
Db	745	AGCCTTGCTGAACGGGGCACCTGCCACCTGATCTGTGCCACCGGACCACCGGTGT 804		
Qy	529	GCTGTCCCAGACAGATTCAAGGGAAATTCGTGAAATAGGTTCTGATGA---CTGCTAT 585		
Db	805	GCCTGCCACCAAGCTTCGCTGGACGGCTCTGCACATCGACCTGATGAGCGTGCCTTC 864		
Qy	586	GTTGGCGATGGCTACTCTTTACCGAGGAAAAATGAATAGACAGTCAAACAGATGCGTGC 845		
Db	865	TTCGGGAACGGCACTGGGTACCGTGGCTGGCGACGACCTCAGCCTCGGCGCCTCAGCTGC 924		
Qy	646	CTTTACTCGAACTCCACACTCTCTTTCAGGAGAATTACAATGTTTAATGAGGATGCT 705		
Db	925	CTGGCTGGAATCCGATCTGCTTACTCAGAGAGCTGCACGTGGACTCGGTGGCGCGCG 984		
Qy	706	GAACCCATGGGATTGGGAAACAAATTTCTCGAUAACCCAGATGCGGACGAAAAGCCC 765		
Db	985	GCCTCTCGGCGTGGCCCCCATGCTACTGCGGAATCCGGAATAACGAGAGGCC 1044		
Qy	766	TGCTGCTTTATTAAGTTTACCAATGACAAGGTGMAATGSGAATACTGTGATCTCTCAGC 825		
Db	1045	TGCTGTACGT---GGTAGAGACAGCGGCTCTCTCTGGAGTACTTCGCCCTTGGAGGCC 1101		
Qy	826	TGCTCAGCCAGGACCTTGCTTACCACAGAGAAAGCCCACTTGAGCATCAACCAAGCTT 885		
Db	1102	TGCGAATCCCT-----CACCAGAGTCCAATGTCACCGATCTCTCTGGCGACCTTG 1152		
Qy	886	CCGGGTTTTGACTCTCTGTGGAAGACTGATAGTAGAGAGGAAGATCAAGAGATCTAT 945		
Db	1153	CCTGAGCCAGCTTCCCGGGGGCCAGGCTCGCGGAGAGGACAAAGAAGAGGAGCTTTC 1212		
Qy	946	GGAGGCTTTAAGAGCAGCGCGGCAAGCACCCATGGCAGGCGTCCCTCCAGTCTCTGCTG 1005		
Db	1213	CTCGGCCACGATCATCTGGCGGCTCCTCTCTCGTCCCGGCTCGCACCCCTG----- 1265		
Qy	1006	CCTCTGACATCTCCATGCCCCNAGGCCACTTCTGTGTGGGGCGTGTATCCACCCTGCG 1065		
Db	1266	--GCTGGCGGCCATCTPACATCGGGGACAGCTTCTGGCGGGAGCGCTGTGTCCACACTGC 1323		
Qy	1066	TGGGTGCTCACTGCTGCCCACTGCACCGACATAAAAAACAGACATCTAA-----AGGTG 1119		
Db	1324	TGGGTGTTGTGGCCGCCCACTCTTCTCCACAGCCCCCCCAGGGACAGCGTCTCCGTTG 1383		
Qy	1120	GTCTAGGGGACAGGACCTGAAGAAAGAAATTCATGACGAGCTTTAGGTGGAG 1179		
Db	1384	GTCTGGGCGACACTTCTTCAACCGACAGCGAGCTGACGACAGCTTCGGCATCTGAG 1443		
Qy	1180	AAGATATTCAAGTACAGGCACCTACAAATGAAGAGATGAGATTCCCCACAATGATATGCA 1239		
Db	1444	AAGTACATCCGTTACACCTGTACTCGGTGTTCAACCCACGCAACCA---GNCTCTGT 1500		
Qy	1240	TTGCTCAAGTTAAAGCCATGATGTTGCTACTGTGCTTAGAATCCAAATACGTGAAGACT 1299		
Db	1501	CTGATCCGCTGAAGAAGAAAGGGGACCGCTGTGCCACACGCTCGCAGTTCTGTGACGCC 1560		
Qy	1300	GTGTGCTTCCTGAT-----GGGTCTTCTTCCCTCTGGGAGTGAGTGCACATCTCTGCG 1353		
Db	1561	ATCTGCTGCCGAGCCCGGAGCACCTTCCCGCNAGGACAAAGTGTCCAGATTGCGGGC 1620		
Qy	1354	TGGGG-----TGTTACAGAAACAGGAAAGGGTCCGCGCAGCTCTCTGATCCCAAGTC 1407		
Db	1621	TGGGGCACTTGATAGAAACGTGAGCGGCTACTCCAGTCTCTGGGAGGCGCTGTGTC 1680		

QY 1408 AAGCTGATTGCCAACACTTTGTGCACTCCGCCCACTCTATGACCAATGATTGATGAC 1467
Db 1681 CCCCTGGTCCCGACACACAAAGTGCAGCAGCCTTACGGTCTACGGCGCGACATCAGCCCC 1740
QY 1468 AGTATGATCTGTGCAGGAATCTTCAGAAACCTTGGGCAAGACACTTGCAGGGTCACTCT 1527
Db 1741 AACATGCTCTGTGCCGGTACTT---CGACTGCAAGTCCGACGCTTGCAGGGGACTCA 1797
QY 1528 GAGGGCCCTGACCTGTGAGAGGACGCGCACTACTCTGTATGGGATAGTGAAGTGG 1587
Db 1798 GGGGGGCCCTCGCCCTGCGAGAAACGCGGTGGCTTACTCTACGGCATCATCACTGG 1857
QY 1588 GGCTCGAGTGTGG-----AAGAGCGCAGGGGTCTACACCAAGTTACCAAAATTCCTG 1641
Db 1858 GGTGACGGCTGCGGGCGGCTCCACAAGCGGGGTCTACACCGCGGTGCGCCCACTATGTG 1917
QY 1642 AATTGGATCAAGCCACCAT 1661
Db 1918 GACTGGATCAACGACCGGAT 1937
RESULT 13
US-09-864-761-27791/c
; Sequence 27791, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aemica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,356
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1

; SEQ ID NO 27791
; LENGTH: 117
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC006037.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.89
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.9
; OTHER INFORMATION: NT HIT: S8182.1 EVALUE 8.00e-60
; OTHER INFORMATION: SWISSPROT HIT: Q25464, EVALUE 1.00e-03
; OTHER INFORMATION: EST_HUMAN HIT: T88666.1, EVALUE 7.00e-55
US-09-864-761-27791
Query Match 7.0%; Score 117; DB 9; Length 117;
Best Local Similarity 100.0%; Pred. No. 4.5e-28;
Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 332 TGCATAATACGTGCAAGGACAAACCCATGTGGCGGGCCCAATGTCTCATTTACCCAGATC 391
Db 117 TGCATAATACGTGCAAGGACAAACCCATGTGGCGGGCCCAATGTCTCATTTACCCAGATC 58
QY 392 CTCCTACTACCGCTGTGTGTAACACCCCTACACAGCTCCAGCTGCTCCCAAG 448
Db 57 CTCCTACTACCGCTGTGTGTAACACCCCTACACAGCTCCAGCTGCTCCCAAG 1
RESULT 14
US-10-087-192-590
; Sequence 590, Application US/10087192
; Publication No. US20020182586A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David W.
; APPLICANT: Engelhard, Eric K.
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: CANCER
; FILE REFERENCE: 529452000122
; CURRENT APPLICATION NUMBER: US/10/087,192
; CURRENT FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: US 09/747,377
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 2059
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 590
; LENGTH: 1302
; TYPE: DNA
; ORGANISM: Mus musculus
US-10-087-192-590
Query Match 5.5%; Score 93.2; DB 13; Length 1302;
Best Local Similarity 49.7%; Pred. No. 1.6e-19;
Matches 586; Conservative 0; Mismatches 523; Indels 59; Gaps 11;
QY 526 TGTGCTGTCCGACCACTTCAAGGGAAATTTCTGTAATAGGT---TCTGATGACTGC 582
Db 154 TGCAGCTGCCCAAGGAAATTCAGGGGAGCACTGTGAGATAGATGCAATCAAAACCTGC 213
QY 583 TATGTTGGGATGGCTACTCTTACCGAGGAAATGATAGACAGTCAACACGATCGC 642
Db 214 TATCATGGAATGGTGTACTTTACCGAGGAAAGGCCAACACTGATACCAAGTGGCCCC 273
QY 643 TGCCTTTTACTGGAATCCCACTCTCTTTGAGGAGAAATTACATGTTTATGAGGAT 702
Db 274 TGCCTGGCTGGAAATGCGCTGCTGCTCTCAGAAACCTACATGCCCCACAGACTGAT 333
QY 703 GCTGAAACCCATGGGATGGGGAACACATTTCTGAGAAACCCAGATGCGGAGGAAAG 762
Db 334 GCTATTAGCTAGGCTGGGGAACACATTTACTGAGAAACCTTGACACACCAAGAGCA 393

763 CCCTGGTCTTTATTAAGATTACCAATGCAAGGTGAAATGGGAATCTGTGATGTCTCA 822
Db CCCTGGTCTATGTCAGATTGGCTAAGCAGTT-----TGTCCTAA 435
Qy GCCTGCTCAGCCAGGAGGTCCTTACCCAGAGGAGAGCCCACTGAGCCATCAACCAAG 882
Db GAATGTCATGGTGCATGAC--TGCTCTCTTACAAAAGAGCTTCTTCTGTGTAGACCAAC 493
Qy CTTCGGGGTTTGACTCTCTGTGAAAGACTGAGATGACAGAGAGAAATCAAGAGATC 942
Db AAGGCTTCCAGTGTGGCC-----AGAAAGCTCTAAGGCCCGCTTTAAGATT 540
Qy TATGAGGCTTTAAGACACGGCGGAGAGCACCACCATGGCAGGAGTCCCTCCAGTCTCTCG 1002
Db GTTGGGGAGAAATTCATGAGGTGAGAACAGCCCTGGTTCGAGCCAT-----CTAC 594
Qy CTGCTCTGACCATCTCCATGCCAGGGGCACCTTCTGTGTGGGGCTGATCCACCC 1062
Db CAGAAGAACAGGAGGAGTCTCTCCCTCTTTAAATGTGTGGAGTCTCATCAGTCT 654
Qy TGCTGGGTGCTCATCTGTGCCCATCTGAC-----GACATAAAACAGACATCTAAG 1116
Db TGCTGGGTGGCAGTGGCGCACCTGCTTCACTCCCAAGAGAAATCAAGTT 714
Qy GTGGTGTAGGGACAGGACCTGAAGAAAGAAATTCATGAGCAGAGCTTTAGGGTG 1176
Db GTCTACCTGGTCACTGAGAGAGCTCTTATATCTCGAGAGATGAAGTTTGGGTG 774
Qy GAGAAGATATTCAAGTACAGCCATCAATGAAGAGATGATATCCCAATATGATTT 1236
Db GAGCAGCTCATCTTGCAAGATATACAGGAGGAGCAGCTGGCTACCATATATGATTT 834
Qy GCATTGCTCAAGTTAAGCCAGTGTGCTACTGTCTAGATCCAAATCAATAGTGAAG 1296
Db GCTTGTCTGAAGATAGTACAGAGGAGGAGGAGTGCACGCCATCCAGTCCATACAG 894
Qy ACTGTGTGCTTCCCTGATGGGTCCTT-----TCCTCTGGAGGTGAGTGCCACATCTCT 1350
Db ACCATCTGCTGCCGCCAAGTTTACTGATGCTCGGTTTGGTTTCAGACTGTGAGATCACT 954
Qy GCCTGGGT--GTTACAGAACAGAGAGGTCGGCCAGCTCTCTGATG--CCAAA 1404
Db GGCTTGGAAAAGAGTCTGAAAAGTGAATCTCTATCCMAAGACCTGAAAATGTCCGTC 1014
Qy GTCAAGCTGATGCCAACACTTTGTGCAACTCCCGCACTCTATGACCACTATGATTCAT 1464
Db GTAAAGCTTGTCTCATGAACAGTGTATGACGCCCACTACTATGGCTCTGAAATTAAT 1074
Qy GACAGTATGATCTGTGAGGAAATCTTCAAGAACCTGGGCAAGACACCTGCCAGGTGAC 1524
Db TATAAATGCTGTGCTCGGACCCAGAGTGGAAAC---AGATTCTGCAAGGGCGAT 1131
Qy TCTGGAGGCCCTGACCTGTGAGAAGGACGGCACCTACTAGTCTTATGGGATGTGAGC 1584
Db TCTGGAGGACCGTTATCTGTAACATCGAAGGCCGCCCACTCTGAGTGGGATTTGAGC 1191
Qy TGGGGCTGAGTGTG-----GGAAGAGCCAGGGGTCTACCCCAAGTTACCAATTC 1638
Db TGGGGCCGAGATGTGCAGAGAAAACAAAGCCCGGTGTCTACAGAGGGTCTCACACTTC 1251
Qy CTGAATTTGATCAAGCACCATCAAAAGTGAAGTGG 1676
Db CTGACTGGATTCAATCCCATTTGGAGAGAGAAAGG 1289

RESULT 15
US-10-665-216-10
; Sequence 10, Application US/10665216
; Publication No. US20040043957A1
; GENERAL INFORMATION:
; APPLICANT: Brenda F. Baker
; APPLICANT: Susan M. Freier

APPLICANT: Andrew T. Watt
; TITLE OF INVENTION: ANTISENSE MODULATION OF UROKINASE PLASMINOGEN ACTIVATOR EXPRESSIC
; FILE REFERENCE: RTS-0188
; CURRENT APPLICATION NUMBER: US/10/665,216
; PRIOR FILING DATE: 2003-09-19
; PRIOR APPLICATION NUMBER: US/09/821,972
; PRIOR FILING DATE: 2001-03-30
; NUMBER OF SEQ ID NOS: 168
; SEQ ID NO 10
; LENGTH: 2299
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (59)....(1360)
US-10-665-216-10

Query Match 5.5%; Score 93.2; DB 13; Length 2299;
Best Local Similarity 49.7%; Pred. No. 2.3e-19;

Matches 586; Conservative 0; Mismatches 523; Indels 69; Gaps 11;

Qy 526 TGTGCTGTCCCGACCCAGTTCAAGGGGAAATTTCTGTGAATAGGT---TCTGATGACTGC 582
Db 212 TGCAGCTGCCAAGGAAATTCAGGGGGAGCACTGTGAGATAGATGCATCAAAAACCTGC 271
Qy 583 TATGTTGGCGATGGCTACTCTTACCGAGGAAATGAATAGCAGCTCAACAGCATGCG 642
Db 272 TATCATGGAATGGTGAATCTTACCGAGGAAAGGCCAACACTGATACCAAGGTGGCCC 331
Qy 643 TGCCTTTTACTGGAATCTCCACCTCTCTTTCAGGAGAAATTACAACATGTTTATGGAGAT 702
Db 332 TGCCTGGCTGGAATGCGCTGCTCTTCAAGAACCTTACAATGCCACAGACCTGAT 391
Qy 703 GCTGAAACCATGGGATTTGGGAAACAATTTCTGCAGAAACCCAGATCGGACGAAAG 762
Db 392 GCTATTAGCTAGGCTGGGAAACAATTTACTGCAGAACCCCTGACAAACAGAAAGCGA 451
Qy 763 CCCTGGTCTTTATTAAGTTACCAATGCAAGGTGAATGGGAATCTGTGATGTCTCA 822
Db 452 CCCTGGTCTATGTGAGATTGGCTTAAGCAGTT-----TGTCCTAA 493
Qy 823 GCCTGTCTCAGCCAGGAGTTCCTTACCAGAGGAAAGCCCACTGAGCCATCAACCAAG 882
Db 494 GAATGATGCTGTGATGAC--TGCTCTCTTAGCAAAAAGCCTTCTCTGTGTAGACCAAC 551
Qy 883 CTTCGGGGTTTGACTCTCTGTGGAAGACTGAGATAGCAGAGAGGAAGATCAAGAGATC 942
Db 552 AAGGCTTCCAGTGTGGCC-----AGAAAGCTCTAAGGCCCGCTTTAAGATT 598
Qy 943 TATGAGGCTTTAAGAGCACGGCGGCAAGCACCCATGGCAGGCGTCCCTCCAGTCTCTCG 1002
Db 599 GTTGGGGAGAAATTCATGAGGTGGAGAACCCCTGTGTTCGAGCCAT-----CTAC 652
Qy 1003 CTGCTCTGACCATCTCTCTGCCCCCAGGCCACTTCTGTGGTGGGGCTGTATCCACCCC 1062
Db 653 CAGAAGAACCAAGGAGAGATCTCTCTCTCTTTAAATGTGTGGAGTCTCATCAGTCT 712
Qy 1063 TGCTGGGTGTCTACTGTGCCCCACTGACCC-----GACATAAAACCAAGATCTTAAG 1116
Db 713 TGCTGGGTGGCCAGTGGCGACACTGTCTTCACTCACTCCAAAGAGAAAGAACTACGTT 772
Qy 1117 GTGTGTGTAGGGGACCGAGGACCTGAAAGAAAGAAATTTATGAGCAGAGCTTTAGGGTG 1176
Db 773 GTCTACCTGGGTGAGTCCGAAGAGAGCTCTCTATAATCTCGAGAGATGAAGTTTGAAGTG 832
Qy 1177 GAGAGATATTCAAGTACAGCCACTTACAATGAAGAGATGAGATTCCCCCAATGATATT 1236
Db 833 GAGCAGCTCATCTTGCAGCAATATCTACAGGAGACAGCCTGGCTACCATATGATATT 892
Qy 1237 GCATTGCTCAAGTTAAAGCCAGTGGATGCTGCTGTGCTGTAGAAATCAAAATAGTGAAG 1296
Db 893 GCCTTGTGTAAGATACGTACCAAGCAGGGCCAAATGTGCACAGCCATCAGGTCATACAG 952

Qy	1297	ACTGTGTGCTGCTGATGGGTCCTT-----TCCCTCTGGGAGTGAGTGCCACATCTCT	1350
Db	953	ACCATCTGCTGCCCCCAAGTTTACTGATGCTCCGTTGGTTTCAGACTGTGAGATCACT	1012
Qy	1351	GGCTGGGGT---CTTACAGAAACAGGAAAGGGTCCCGCAGCTCCTGGATG---CCAAA	1404
Db	1013	GGTTTGGAAAGAGTCTGAAAGTGACTTCTCTATCCAAAGAACCTGAAATGTCGTC	1072
Qy	1405	GTCAAGCTGATGCCAACACTTTGTGCAACTCCCGCAACTCTATGACCCACATGATTGAT	1464
Db	1073	GTAAGCTTGTCTCATGAACAGTGTATGCAGCCCCCACTACTATGGTCTGAAATTAAT	1132
Qy	1465	GCAGTATGATCTGTGCAGGAATCTTCAGAAACCTGGGCAAGACACCTGCCAGGGTGAC	1524
Db	1133	TATAAATGCTGTGCTGCGACCCAGAGTGAATAAC---AGATTCTTGCAGGGCGAT	1189
Qy	1525	TCTGGAGCCCCCTGACCTGTGAGAGGACGGCACCTACTACTGCTATGGGATAGTAC	1584
Db	1190	TCTGGAGGACCGCTTATCTGTAACATCGAAGGCCGCCCAACTCTGAGTGGGATTGTGAGC	1249
Qy	1585	TGGGGCTGGAGTGTG-----GGAAGAGCCAGGGGTCTACACCCCAAGTTACCAATTC	1638
Db	1250	TGGGGCCGAGGATGTGAGAGAAACAGCCCGGTGTCTACACGAGGGTCTCACACTTC	1309
Qy	1639	CTGAATTGGATCAAAGCCACCATCAAAAGTGAAAGTGG	1676
Db	1310	CTGACTGGATTCAATCCCAATTGGAGAGAGAAAGG	1347

Search completed: May 25, 2004, 13:22:31
Job time : 538 secs

score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 25, 2004, 04:22:47 ; Search time 4479.5 Seconds
(without alignments)
16284.480 Million cell updates/sec

Title: US-09-912-559-2

Perfect score: 1683

Sequence: 1 atgttgcagatgtctga.....aaagtgaagtgtctctaa 1683

Scoring table:

IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

- 1: gb_ba.*
- 2: gb_bt.*
- 3: gb_in.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vi.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_ma.*
- 20: em_om.*
- 21: em_or.*
- 22: em_ov.*
- 23: em_pat.*
- 24: em_ph.*
- 25: em_pl.*
- 26: em_ro.*
- 27: em_sts.*
- 28: em_un.*
- 29: em_vi.*
- 30: em_htg_hum.*
- 31: em_htg_inv.*
- 32: em_htg_other.*
- 33: em_htg_mus.*
- 34: em_htg_pin.*
- 35: em_htg_rtd.*
- 36: em_htg_mam.*
- 37: em_htg_vrt.*
- 38: em_sv.*
- 39: em_higo_hum.*
- 40: em_higo_mus.*
- 41: em_higo_other.*

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1683	100.0	1683	6	AX383955	AX383955 Sequence
2	1683	100.0	1683	6	BD177603	BD177603 Detection
3	1679.8	99.8	1683	6	AX383954	AX383954 Sequence
4	1679.8	99.8	1683	6	BD177602	BD177602 Detection
5	1679.8	99.8	2408	9	S83182	S83182 hyaluronan-
6	1679.8	99.8	3008	6	AX409021	AX409021 Sequence
7	1679.8	99.8	3008	9	HUMHGFAL	D49742 Human mRNA
8	1678.2	99.7	2251	9	BC031412	BC031412 Homo sapi
9	1141	67.8	2190	10	AK128915	AK128915 Mus muscu
10	1047	62.2	2119	10	BC031775	BC031775 Mus muscu
11	261.2	15.5	188349	9	AL390197	AL390197 Human DNA
12	258	15.3	157750	2	AC006097	AC006097 Homo sapi
13	241	14.3	397	6	AX305372	AX305372 Sequence
14	167.4	9.9	204214	2	AC115771	AC115771 Mus muscu
15	161	9.6	224208	2	AC106236	AC106236 Rattus no
16	161	9.6	249249	2	AC131859	AC131859 Rattus no
17	129.8	7.7	2035	10	AF099017	AF099017 Mus muscu
18	129.8	7.7	2063	10	AF224724	AF224724 Mus muscu
19	128.2	7.6	2067	10	BC019376	BC019376 Mus muscu
20	114.4	6.8	2033	6	E07591	E07591 DNA encodin
21	114.4	6.8	2033	6	I15469	I15469 Sequence 14
22	114.4	6.8	2033	6	I69315	I69315 Sequence 14
23	114.4	6.8	2036	6	AX333070	AX333070 Sequence
24	114.4	6.8	2036	6	AX408965	AX408965 Sequence
25	114.4	6.8	2036	9	HUMHGFAP	D14012 Homo sapien
26	99.8	5.9	2355	4	BTUPA	X85801 B.taurus mR
27	99.8	5.9	2355	4	BOVUKPA	L03546 Bos taurus
28	99	5.9	2375	4	SSUPAR	X02724 Porcine mRN
29	94.6	5.6	1454	10	RNUPAM	X65651 R.norvegicu
30	94.6	5.6	2366	10	RNUTPLAC	X63434 R.norvegicu
31	93.2	5.5	2299	10	MMURKR	X02389 Mouse mRNa
32	93	5.5	970	6	E07615	E07615 DNA encodin
33	93	5.5	970	6	I15462	I15462 Sequence 3
34	93	5.5	970	6	I69308	I69308 Sequence 3
35	87.6	5.2	2245	4	VMPAAIA	M63987 Desmodus ro
36	87.6	5.2	2252	6	A07242	A07242 Artificial
37	87.6	5.2	2252	6	A07250	A07250 Artificial
38	87.6	5.2	2266	4	VMPAAIS	M63986 Desmodus ro
39	87.4	5.2	2512	6	Al9618	Al9618 Mammalian c
40	87.4	5.2	2585	10	BC061565	BC061565 Rattus no
41	87.2	5.2	2445	10	RATPATISS	M23697 Rat tissue-
42	84.8	5.0	2130	4	VMPAPAB	M63989 Desmodus ro
43	84.8	5.0	2137	6	A07246	A07246 Artificial
44	81.2	4.8	1620	6	I08184	I08184 Sequence 4
45	81.2	4.8	1725	4	VMPTPA	J05082 Vampire bat

ALIGNMENTS

RESULT 1	AX383955	AX383955	1683 bp	DNA	linear	PAT 19-MAR-2002
LOCUS	Sequence 2 from Patent EP1182258.					
DEFINITION	AX383955					
ACCESSION	AX383955					
VERSION	AX383955.1	GI:19577526				
KEYWORDS						
SOURCE	Homo sapiens (human)					
ORGANISM	Homo sapiens					
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.					
REFERENCE	1					
AUTHORS	Roemisch, J., Stoeck, H.A., Feussner, A., Lang, W., Weimer, T.,					
	Becker, M., Nerlich, C. and Muth-Naumann, G.					
TITLE	Mutants of the factor vii activating protease and methods for their					

Pred. No. is the number of results predicted by chance to have a

detection
 JOURNAL Patent: EP 1182258-A 2 27-FEB-2002;
 Aventis Behring GmbH (DE)
 FEATURES Location/Qualifiers
 source 1..1683
 /organism="Homo sapiens"
 /mol_type="unassigned DNA"
 /db_xref="taxon:9606"

ORIGIN

Query Match 100.0%; Score 1683; DB 6; Length 1683;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1683; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGTTTCCAGGATGCTGATCCAGTTCCTCTCTTAATGCTCTGGTGGGAAGACA 60
 Db 1 ATGTTTCCAGGATGCTGATCCAGTTCCTCTCTTAATGCTCTGGTGGGAAGACA 60

QY 61 GCCTGTGGGTTCTCCCTGATGTCTTTATTGGAAGCCTGGACCCAGACTGGACCCCTGAC 120
 Db 61 GCCTGTGGGTTCTCCCTGATGTCTTTATTGGAAGCCTGGACCCAGACTGGACCCCTGAC 120

QY 121 CAGTATGATTACAGTACGAGGATTATATCAGAGAGAGACACACAGTACGACATTAC 180
 Db 121 CAGTATGATTACAGTACGAGGATTATATCAGAGAGAGACACACAGTACGACATTAC 180

QY 181 CATGCTCAGAAATCCTGACTGGTACTACACTGAGGACCAAGCTGATCCATGCCAGCCCAAC 240
 Db 181 CATGCTCAGAAATCCTGACTGGTACTACACTGAGGACCAAGCTGATCCATGCCAGCCCAAC 240

QY 241 CCCTGTGAACACGGTGGGAGCTGCTGTCATCGGAGACCTTCATGACAGCTGCCCTG 300
 Db 241 CCCTGTGAACACGGTGGGAGCTGCTGTCATCGGAGACCTTCATGACAGCTGCCCTG 300

QY 301 GCTCTCTTCTCTGGAAATAGTGTGAGAAAGTGCAGAAATGCAAAATACGTCGAAAGCAACCCATGT 360
 Db 301 GCTCTCTTCTCTGGAAATAGTGTGAGAAATGCAAAATACGTCGAAAGCAACCCATGT 360

QY 361 GGCCGGGGCAATGTCTCATTACCCAGAGTCTCCCTTACCTACCGCTGTGCTGTAAACAC 420
 Db 361 GGCCGGGGCAATGTCTCATTACCCAGAGTCTCCCTTACCTACCGCTGTGCTGTAAACAC 420

QY 421 CCTTACACAGTCCACGCTGCTCCCAAGTGTGTTCTGTATGAGGCAAAACCCCTGCCAG 480
 Db 421 CCTTACACAGTCCACGCTGCTCCCAAGTGTGTTCTGTATGAGGCAAAACCCCTGCCAG 480

QY 481 AATGGGCTACCTGCTCCCGCATAGCGGAGATCCAAAGTTCACCTGTGCTGTCCCGAC 540
 Db 481 AATGGGCTACCTGCTCCCGCATAGCGGAGATCCAAAGTTCACCTGTGCTGTCCCGAC 540

QY 541 CAGTTCAAGGGAAATTCGTGAAATAGGTTCTGTATGATGCTGCTATGTCGCGATGCTAC 600
 Db 541 CAGTTCAAGGGAAATTCGTGAAATAGGTTCTGTATGATGCTGCTATGTCGCGATGCTAC 600

QY 601 TCTTACCGAGGAAATGATAGACAGTCAACACAGATCGTGCCTTTACTGGAACTCC 660
 Db 601 TCTTACCGAGGAAATGATAGACAGTCAACACAGATCGTGCCTTTACTGGAACTCC 660

QY 661 CACCTCTCTGAGGAGATTAACATGTTTATGAGGATGCTGAAACCCATGGATT 720
 Db 661 CACCTCTCTGAGGAGATTAACATGTTTATGAGGATGCTGAAACCCATGGATT 720

QY 721 GGGGAACACAATTTCTGCAGAAACCCAGATGCGGAGGAAAGCCCTGGTCTTTATAA 780
 Db 721 GGGGAACACAATTTCTGCAGAAACCCAGATGCGGAGGAAAGCCCTGGTCTTTATAA 780

QY 781 GTTACCAATGACAGGTAATGGGATACGTGATGCTCAGCCTGCTCAGCCAGGAC 840
 Db 781 GTTACCAATGACAGGTAATGGGATACGTGATGCTCAGCCTGCTCAGCCAGGAC 840

QY 841 GTTGCTTACCCAGAGGAAAGCCCACTGAGCCATCAACCAAGTTCGCGGTTTGACTCC 900
 Db 841 GTTGCTTACCCAGAGGAAAGCCCACTGAGCCATCAACCAAGTTCGCGGTTTGACTCC 900

QY 901 TGTGAAAGACTGAGATAGCAGAGAGAGAGATCAAGAGAAATCTATGAGAGCTTTAAGAGC 960
 Db 901 TGTGAAAGACTGAGATAGCAGAGAGAGAGATCAAGAGAAATCTATGAGAGCTTTAAGAGC 960

QY 961 ACGGCGGCAAGCACCACATGGCAGCGTCCCTCCAGTCCCTGCTGCTGCTGCTGCTGCTGCT 1020
 Db 961 ACGGCGGCAAGCACCACATGGCAGCGTCCCTCCAGTCCCTGCTGCTGCTGCTGCTGCTGCT 1020

QY 1021 ATGCCCCAGGCCACTTCTGTGTGGGCGCTGATCCACCCCTGCTGGTGGTCTGCTGCT 1080
 Db 1021 ATGCCCCAGGCCACTTCTGTGTGGGCGCTGATCCACCCCTGCTGGTGGTCTGCTGCT 1080

QY 1081 GCCCACTGCACCCGACATATAAACCAGACATCTAAAGTGTGTAGGGGACCGAGCCTG 1140
 Db 1081 GCCCACTGCACCCGACATATAAACCAGACATCTAAAGTGTGTAGGGGACCGAGCCTG 1140

QY 1141 AAGAAAGAGAAATTTCAATGACAGAGCTTTAGGGTGCAGAGATATTTCAAGTACAGCCAC 1200
 Db 1141 AAGAAAGAGAAATTTCAATGACAGAGCTTTAGGGTGCAGAGATATTTCAAGTACAGCCAC 1200

QY 1201 TACATGAAGAGATGAGATTCCCAATGATATTCGATTTGCTCAAGTTAAAGCCAGTG 1260
 Db 1201 TACATGAAGAGATGAGATTCCCAATGATATTCGATTTGCTCAAGTTAAAGCCAGTG 1260

QY 1261 GATGGTCACTGTGCTCTAGAAATCCAAATACGTGAAGACTGTGTGCTTGCCTGATGGTCC 1320
 Db 1261 GATGGTCACTGTGCTCTAGAAATCCAAATACGTGAAGACTGTGTGCTTGCCTGATGGTCC 1320

QY 1321 TTTCCCTCTGGAGTGTGCTGACATCTCTGGCTGGGGTGTACAGAAAACAGGAAAGGG 1380
 Db 1321 TTTCCCTCTGGAGTGTGCTGACATCTCTGGCTGGGGTGTACAGAAAACAGGAAAGGG 1380

QY 1381 TCCCGCCAGCTCTCTGGATGTCAGAAAGTCAAGCTGATTCGCAACACTTTGTGCAACTCCCGC 1440
 Db 1381 TCCCGCCAGCTCTCTGGATGTCAGAAAGTCAAGCTGATTCGCAACACTTTGTGCAACTCCCGC 1440

QY 1441 CAATCTATGACCACTGATGATGACAGTATGATCTGTGCGAGGAATCTTCAGAAACCT 1500
 Db 1441 CAATCTATGACCACTGATGATGACAGTATGATCTGTGCGAGGAATCTTCAGAAACCT 1500

QY 1501 GGGCAAGACACTGCCAGGCTGACTCTGGAGGCCCCCTGACCTGTGAGAGGACGGCACC 1560
 Db 1501 GGGCAAGACACTGCCAGGCTGACTCTGGAGGCCCCCTGACCTGTGAGAGGACGGCACC 1560

QY 1561 TACTACGCTATGAGATGAGTGTGGGCTGTGGAGTGTGAGAGAGGCGGAGCTTAC 1620
 Db 1561 TACTACGCTATGAGATGAGTGTGGGCTGTGGAGTGTGAGAGAGGCGGAGCTTAC 1620

QY 1621 ACCCAAGTTACCAAAATTCCTGAATTGGATCAAGCCCAACCATCAAAAGTGAAGTGGCTTC 1680
 Db 1621 ACCCAAGTTACCAAAATTCCTGAATTGGATCAAGCCCAACCATCAAAAGTGAAGTGGCTTC 1680

QY 1681 TAA 1683
 Db 1681 TAA 1683

RESULT 2
 BD177603
 LOCUS
 DEFINITION Detection method with the use of factor VII activating protease
 mutant and specific antibody.
 ACCESSION BD177603
 VERSION BD177603.1 GI:30014865
 KEYWORDS JP 2002291485-A/2.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 1683)
 Roemisch, J., Stoeck, H.A., Feussner, A., Lang, W., Weimer, T.,
 Becker, M., Nerlich, C. and Naumann, G.M.

TITLE Detection method with the use of factor VII activating protease
 JOURNAL Patent: JP 2002291486-A 2 08-OCT-2002;
 COMMENT AVENTIS BEHRING GMBH
 OS Homo sapiens (human)
 PN JP 2002291486-A/2
 PD 08-OCT-2002
 PF 25-JUL-2001 JP 2001224423
 PR 26-JUL-2000 DE 10036641:4,10-OCT-2000 DE 10050040:4 PR
 21-OCT-2000 DE 10052319:6,12-APR-2001 DE 10118706:8 PI
 JUEGEN ROEMISCH, HANS ARNOLD STOEHR, ANNETTE FEUSSNER, WIEGAND
 LANG,
 PI THOMAS WEIMER, MARGRET BECKER, CLAUDIA NERLICH, GUDRUN MUTH PI
 NAUMANN
 PC C12N15/09, C07K16/40, C12N9/50, C12Q1/68, G01N33/53, G01N33/533, PC
 G01N33/534
 PC G01N33/535//C12P21/08, C12N15/00
 CC Detection method with the use of factor VII activating CC
 protease mutant and
 CC specific antibody
 FH Key Location/Qualifiers
 FT source 1..1683 /organism='Homo sapiens (human)'.
 FT Location/Qualifiers
 1..1683
 /organism='Homo sapiens'
 /mol_type='genomic DNA'
 /db_xref='taxon:9606'
 ORIGIN
 Query Match 100.0%; Score 1683; DB 6; Length 1683;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1683; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 ATGTTTGCAGAGTCTGATCTCCATGTTCTGCTGTTAATGCTCTGTTGGGAAAGACA 60
 DB 1 ATGTTTGCAGAGTCTGATCTCCATGTTCTGCTGTTAATGCTCTGTTGGGAAAGACA 60
 QY 61 GCCTGTGGTCTCCCTGATGCTTTATTGGAAGCCCTGGACCCAGACTGACCCCTGAC 120
 DB 61 GCCTGTGGTCTCCCTGATGCTTTATTGGAAGCCCTGGACCCAGACTGACCCCTGAC 120
 QY 121 CAGTATGATTACAGTACAGAGTATATAATCAGGAAGAACACAGTAGCACACTTACC 180
 DB 121 CAGTATGATTACAGTACAGAGTATATAATCAGGAAGAACACACAGTAGCACACTTACC 180
 QY 181 CATGCTGAGATCTCTGATCTGTACTACACTGAGGACCAAGCTGATCCATGCCGCCAAC 240
 DB 181 CATGCTGAGATCTCTGATCTGTACTACACTGAGGACCAAGCTGATCCATGCCGCCAAC 240
 QY 241 CCCTGTGAACACGGTGGGACTGCTCGTCCATGGGACACCTTCACATGCACTGCTG 300
 DB 241 CCCTGTGAACACGGTGGGACTGCTCGTCCATGGGACACCTTCACATGCACTGCTG 300
 QY 301 GCTCCTTTCTCTGGGAATAGTGTGAGAAAGTGCAAAATACGTGCAAGGACCAACCCATGT 360
 DB 301 GCTCCTTTCTCTGGGAATAGTGTGAGAAAGTGCAAAATACGTGCAAGGACCAACCCATGT 360
 QY 361 GGCCGGGCCAATGCTCTATTACCCAGAGTCTCCCTACTACCGCTGTGTGTAACAC 420
 DB 361 GGCCGGGCCAATGCTCTATTACCCAGAGTCTCCCTACTACCGCTGTGTGTAACAC 420
 QY 421 CTTTACAGGTCACAGTGTCTCCCAAGTGGTTCCTGTATGACGCGCAAAACCCCTGCGAG 480
 DB 421 CTTTACAGGTCACAGTGTCTCCCAAGTGGTTCCTGTATGACGCGCAAAACCCCTGCGAG 480
 QY 481 AATGGGGTACTGCTCCGGGATTAAGCGGAGATCCAGTTTCACTGTGCTGTCCCGAC 540
 DB 481 AATGGGGTACTGCTCCGGGATTAAGCGGAGATCCAGTTTCACTGTGCTGTCCCGAC 540
 QY 541 CAGTTCAAGGGGAAATTCCTGTAATAGTCTGATGATGCTGATGTTGGGATGGGTAC 600
 DB 541 CAGTTCAAGGGGAAATTCCTGTAATAGTCTGATGATGCTGATGTTGGGATGGGTAC 600

601 TCTTACCAGGAGAAATGAATAGGACAGTCAACAGCATGGTGCCTTTACTGGAACTCC 660
 DB 601 TCTTACCAGGAGAAATGAATAGGACAGTCAACAGCATGGTGCCTTTACTGGAACTCC 660
 QY 661 CACCTCCTCTTTCAGGAGAAATTAACAACATGTTTATGAGGATGCTGAAACCCATGGGATT 720
 DB 661 CACCTCCTCTTTCAGGAGAAATTAACAACATGTTTATGAGGATGCTGAAACCCATGGGATT 720
 QY 721 GGGGAACACATTTCTGAGAAACCCAGATCGGACGAAAGCCCTGCTGTTATTAAA 780
 DB 721 GGGGAACACATTTCTGAGAAACCCAGATCGGACGAAAGCCCTGCTGTTATTAAA 780
 QY 781 GTTACCAATGACAAAGTGAATGGAAATCTGTGATGCTCAGCCTGCTCAGCCAGGAC 840
 DB 781 GTTACCAATGACAAAGTGAATGGAAATCTGTGATGCTCAGCCTGCTCAGCCAGGAC 840
 QY 841 GTTGCTTACCCAGAGAAAGCCCACTGAGCCATCAACCAAGCTTCGGGGTTTACTCC 900
 DB 841 GTTGCTTACCCAGAGAAAGCCCACTGAGCCATCAACCAAGCTTCGGGGTTTACTCC 900
 QY 901 TGTGAAAGACTGAGTAGCAGAGAGAGATCAAGAGATCTATGAGGCTTTTAAGAGC 960
 DB 901 TGTGAAAGACTGAGTAGCAGAGAGAGATCAAGAGATCTATGAGGCTTTTAAGAGC 960
 QY 961 ACGGGGGCAAGCACCCTGAGGAGGCTCCCTCCAGTCTCTGCTGCTGACCATCTCC 1020
 DB 961 ACGGGGGCAAGCACCCTGAGGAGGCTCCCTCCAGTCTCTGCTGCTGACCATCTCC 1020
 QY 1021 ATGCCCCAGGGCCACTTCTGTGGTGGGCGCTGATCCACCCCTGCTGGTGTCTACTGCT 1080
 DB 1021 ATGCCCCAGGGCCACTTCTGTGGTGGGCGCTGATCCACCCCTGCTGGTGTCTACTGCT 1080
 QY 1081 GCCCACTGCACCGACATAAAAAACCAAGATCTTAAAGTGGTGTCTAGGGGACAGGACCTG 1140
 DB 1081 GCCCACTGCACCGACATAAAAAACCAAGATCTTAAAGTGGTGTCTAGGGGACAGGACCTG 1140
 QY 1141 AAGAAAGAGATTTTCATGAGCAGAGCTTTAGGGTGCAGAGATATTCAGTACAGCCAC 1200
 DB 1141 AAGAAAGAGATTTTCATGAGCAGAGCTTTAGGGTGCAGAGATATTCAGTACAGCCAC 1200
 QY 1201 TACAATGAAGAGATGAGATTTCCCAACATGATATTGCAATGCTCAAGTTAAAGCCAGTG 1260
 DB 1201 TACAATGAAGAGATGAGATTTCCCAACATGATATTGCAATGCTCAAGTTAAAGCCAGTG 1260
 QY 1261 GATGTCCTGCTCTAGAAATCCAAATACGTGAAGACTGTGCTTGGCTGATGGTCC 1320
 DB 1261 GATGTCCTGCTCTAGAAATCCAAATACGTGAAGACTGTGCTTGGCTGATGGTCC 1320
 QY 1321 TTTCCCTCTGGGAGTGAAGTCCACATCTCTGGCTGGGGTGTTCAGAAACAGGAAAGGG 1380
 DB 1321 TTTCCCTCTGGGAGTGAAGTCCACATCTCTGGCTGGGGTGTTCAGAAACAGGAAAGGG 1380
 QY 1381 TCCCGCCAGCTCTGTGATGCCAAAGTCAAGCTGATTCGCAACACTTTGTGCAACTCCCGC 1440
 DB 1381 TCCCGCCAGCTCTGTGATGCCAAAGTCAAGCTGATTCGCAACACTTTGTGCAACTCCCGC 1440
 QY 1441 CAACTCTATGACCAACATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1500
 DB 1441 CAACTCTATGACCAACATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1500
 QY 1501 GGGCAAGACACCTGCCAGGGTGAATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1560
 DB 1501 GGGCAAGACACCTGCCAGGGTGAATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1560
 QY 1561 TACTAGCTCTATGGGATGAGTGGGCTGGAGTGGAGAGGAGGAGGAGGAGGAGGAGGAGGAG 1620
 DB 1561 TACTAGCTCTATGGGATGAGTGGGCTGGAGTGGAGAGGAGGAGGAGGAGGAGGAGGAGGAG 1620
 QY 1621 ACCCAAGTTACCAAAATTCCTGAATTTGGATTCAAAGCCACCATCAAAAGTGAAGTGGCTTC 1680
 DB 1621 ACCCAAGTTACCAAAATTCCTGAATTTGGATTCAAAGCCACCATCAAAAGTGAAGTGGCTTC 1680

QY 1681 TAA 1683
Db 1681 TAA 1683

RESULT 3
AX383954
LOCUS AX383954 1683 bp DNA linear PAT 19-MAR-2002
DEFINITION Sequence 1 from Patent EP1182258.
ACCESSION AX383954
VERSION AX383954.1 GI:19577525
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
1 Roemisch, J., Stoeck, H.A., Feussner, A., Lang, W., Weimer, T.,
Becker, M., Nerlich, C. and Much-Naumann, G.
Mutants of the factor VII activating protease and methods for their
detection

JOURNAL Patent: EP 1182258-A 1 27-FEB-2002;
Aventis Behring GmbH (DE)

FEATURES
source
1..1683
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN
Query Match 99.8%; Score 1679.8; DB 6; Length 1683;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 168; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATGTTTGCCAGGATGTCATCTCCATGTTCTCTCTCTTAATGGCTCTGTGGGAAAGACA 60
Db 1 ATGTTTGCCAGGATGTCATCTCCATGTTCTCTCTCTTAATGGCTCTGTGGGAAAGACA 60

QY 61 GCCTGTGGGTCTCCCTGATGTTCTTTATGGAAAGCTGGACCCAGACTGGACCCCTGAC 120
Db 61 GCCTGTGGGTCTCCCTGATGTTCTTTATGGAAAGCTGGACCCAGACTGGACCCCTGAC 120

QY 121 CAGTATGATTACAGTACGAGGATTAATACAGGAGAGAACACCAAGTACGACATTACC 180
Db 121 CAGTATGATTACAGTACGAGGATTAATACAGGAGAGAACACCAAGTACGACATTACC 180

QY 181 CATGCTGAGAACTCTGACTGGTACTACATGAGGACCAAGCTGATCCATGCCAGCCCAAC 240
Db 181 CATGCTGAGAACTCTGACTGGTACTACATGAGGACCAAGCTGATCCATGCCAGCCCAAC 240

QY 241 CCCTGTCAACACGGTGGGAGCTGCTGCTCCATGGGAGCACCTTCAATGACGCTGCCTG 300
Db 241 CCCTGTCAACACGGTGGGAGCTGCTGCTCCATGGGAGCACCTTCAATGACGCTGCCTG 300

QY 301 GCTCCTTTCTCTGGGAATAAGTGTGAGAAAGTGAGAAATACGTGGAAGCAACCCATGT 360
Db 301 GCTCCTTTCTCTGGGAATAAGTGTGAGAAAGTGAGAAATACGTGGAAGCAACCCATGT 360

QY 361 GGCCGGGGCCAAATGTCATTACCCAGAGTCTCCCTACTACCGCTGTGTCTGTAACAC 420
Db 361 GGCCGGGGCCAAATGTCATTACCCAGAGTCTCCCTACTACCGCTGTGTCTGTAACAC 420

QY 421 CTTTACACAGTCCCAAGTGTCTCCCAAGTGGTTCCTGTATGAGAGCCAAACCCCTGCCAG 480
Db 421 CTTTACACAGTCCCAAGTGTCTCCCAAGTGGTTCCTGTATGAGAGCCAAACCCCTGCCAG 480

QY 481 AATGGGCTACTCTGCTCCCGCAATAGCGGAGATCCAAAGTTCACCTGTGCTGCCGAC 540
Db 481 AATGGGCTACTCTGCTCCCGCAATAGCGGAGATCCAAAGTTCACCTGTGCTGCCGAC 540

QY 541 CAGTTCAAGGGGAAATCTGTGAATAGTTCATGATGACTGCTATGTTGCCGATGGCTAC 600
Db 541 CAGTTCAAGGGGAAATCTGTGAATAGTTCATGATGACTGCTATGTTGCCGATGGCTAC 600

QY 601 TCTTACCGAGGAAATGAAATAGACAGTCAACACAGCATCGTGCTCTTTACTGAACTCC 660
Db 601 TCTTACCGAGGAAATGAAATAGACAGTCAACACAGCATCGTGCTCTTTACTGAACTCC 660

QY 661 CACCTCCTCTTGCAGGAGAAATTACACATGTTTATGAGGATGCTGAAACCCATGGGATT 720
Db 661 CACCTCCTCTTGCAGGAGAAATTACACATGTTTATGAGGATGCTGAAACCCATGGGATT 720

QY 721 GGGGAACACAAATTTCTGCAGAAACCCAGATCGGAGAGAAAGCCCTGGTCTTTATTA 780
Db 721 GGGGAACACAAATTTCTGCAGAAACCCAGATCGGAGAGAAAGCCCTGGTCTTTATTA 780

QY 781 GTTACCAATGACAAAGTGAATGGAATACTGTGATGCTCTCAGCTGCTCAGGCCAGGAC 840
Db 781 GTTACCAATGACAAAGTGAATGGAATACTGTGATGCTCTCAGCTGCTCAGGCCAGGAC 840

QY 841 GTTCCCTACCCAGAGGAAAGCCCACTGAGCCATCAACCAAGCTTCGGGGTTTGACTCC 900
Db 841 GTTCCCTACCCAGAGGAAAGCCCACTGAGCCATCAACCAAGCTTCGGGGTTTGACTCC 900

QY 901 TGTGAAAGACTGAGATAGCAGAGAGAGATCAAGAGAAATCTATGAGGCTTTAAGAGC 960
Db 901 TGTGAAAGACTGAGATAGCAGAGAGAGATCAAGAGAAATCTATGAGGCTTTAAGAGC 960

QY 961 ACGGGGGCAAGCACTGATGGAGGCTCCCTCAGTCCCTCGCTGCTGCTGCAATCTCC 1020
Db 961 ACGGGGGCAAGCACTGATGGAGGCTCCCTCAGTCCCTCGCTGCTGCTGCAATCTCC 1020

QY 1021 ATGCCCCCAGGGCCACTTCTGTGTGGGGCGCTGATCCACCCCTGCTGGGTCTCACTGCT 1080
Db 1021 ATGCCCCCAGGGCCACTTCTGTGTGGGGCGCTGATCCACCCCTGCTGGGTCTCACTGCT 1080

QY 1081 GCCCACTGCAACGATTAATAAACCAGACATCTAAGGTGTGTGAGGAGCCAGGACCTG 1140
Db 1081 GCCCACTGCAACGATTAATAAACCAGACATCTAAGGTGTGTGAGGAGCCAGGACCTG 1140

QY 1141 AAGAAAGAGAAATTTCAATGAGCAGAGCTTTAGGTGAGAGAGATATTCAAGTACAGCCAC 1200
Db 1141 AAGAAAGAGAAATTTCAATGAGCAGAGCTTTAGGTGAGAGAGATATTCAAGTACAGCCAC 1200

QY 1201 TACAATGAAGAGATGAGATTTCCCAATGATTTGCAATGATTTGCAATGTTAAAGCCAGTG 1260
Db 1201 TACAATGAAGAGATGAGATTTCCCAATGATTTGCAATGATTTGCAATGTTAAAGCCAGTG 1260

QY 1261 GATGTCATGCTGTCTAGAAATCCAAATACGTGAAGACTGTGTGCTTGTGCTGATGGGTCC 1320
Db 1261 GATGTCATGCTGTCTAGAAATCCAAATACGTGAAGACTGTGTGCTTGTGCTGATGGGTCC 1320

QY 1321 TTTCCCTCTGGGAGTGAAGTCCCAATGATTTGCAATGATTTGCAATGTTAAAGCCAGTG 1380
Db 1321 TTTCCCTCTGGGAGTGAAGTCCCAATGATTTGCAATGATTTGCAATGTTAAAGCCAGTG 1380

QY 1381 TCCCGCCAGCTCTGGATGCCAAAGTCAAGTGAATGCCAACACTTTGTGCAACTCCCGC 1440
Db 1381 TCCCGCCAGCTCTGGATGCCAAAGTCAAGTGAATGCCAACACTTTGTGCAACTCCCGC 1440

QY 1441 CAACCTCTATGACCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1500
Db 1441 CAACCTCTATGACCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1500

QY 1501 GGGCAAGACACTGCGAGGGTGAATCTGAGGGCCCCCTGACCTGTGAGAGAGCGGACCC 1560
Db 1501 GGGCAAGACACTGCGAGGGTGAATCTGAGGGCCCCCTGACCTGTGAGAGAGCGGACCC 1560

QY 1561 TACTACGTCTATGGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1620
Db 1561 TACTACGTCTATGGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1620

QY 1621 ACCCAAGTACCAATTTCTGAAATGATCAACCCAGCATCAAAAGTGAAGTGGGCTTC 1680
Db 1621 ACCCAAGTACCAATTTCTGAAATGATCAACCCAGCATCAAAAGTGAAGTGGGCTTC 1680

QY	1681	TAA 1683
Db	1681	TAA 1683
RESULT 4		
BD177602		
LOCUS	1683 bp DNA linear PAT 16-APR-2003	
DEFINITION	Detection method with the use of factor VII activating protease mutant and specific antibody.	
ACCESSION	BD177602	
VERSION	BD177602.1 GI:30014864	
KEYWORDS	JP 2002291486-A/1.	
SOURCE	Homo sapiens (human)	
ORGANISM	Homo sapiens	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.	
AUTHORS	Roemisch,J., Stoehr,H.A., Feussner,A., Lang,W., Weimer,T., Becker,M., Nerlich,C. and Naumann,G.M.	
TITLE	Detection method with the use of factor VII activating protease mutant and specific antibody	
JOURNAL	Patent: JP 2002291486-A 1 08-OCT-2002;	
COMMENT	AVENTIS BEHRING GMBH OS Homo sapiens (human) PN JP 2002291486-A/1 PD 08-OCT-2002 PF 25-JUN-2001 JP 2001224423 PR 26-JUL-2000 DE 10036641:4,10-OCT-2000 DE 10050040:4 PR 21-OCT-2000 DE 10052319:6,12-APR-2001 DE 10118706:9 PI JUEGEN ROEMISCH,HANS ARNOLD STOEHR,ANNETTE FEUSSNER,WIEGAND PI PI THOMAS WEIMER,MARGRET BECKER,CLAUDIA NERLICH,GUDRUN MUTH PI NAUMANN PC C12N15/09,C07K15/40,C12N9/50,C12Q1/68,G01N33/53,G01N33/533, PC GOIN33/534, FC GOIN33/535//C12P21/08,C12N15/00 CC Detection method with the use of factor VII activating CC protease mutant and CC specific antibody FH key Location/Qualifiers FT source i.1683 /organism='Homo sapiens (human)'. FT location/Qualifiers 1..1683 /mol_type='genomic DNA' /db_xref='taxon:9606'	
FEATURES		
source		
ORIGIN		
Query Match 99.8%; Score 1679.8; DB 6; Length 1683; Best Local Similarity 99.9%; Pred. No. 0; Matches 1681; Conservative 0; Mismatches 2; Indels 0; Gaps 0;		
QY	1	ATGTTTGCCAGGATGTCGATCCATCGTCTGTCTGTTAATGGCTCTGGTGGGAAGACA 60
Db	1	ATGTTTGCCAGGATGTCGATCCATCGTCTGTCTGTTAATGGCTCTGGTGGGAAGACA 60
QY	61	GCTGTGGGTTCTCCCTGATGCTTTATTGGAAGCCTGGACCAGACTGGACCCCTGAC 120
Db	61	GCTGTGGGTTCTCCCTGATGCTTTATTGGAAGCCTGGACCAGACTGGACCCCTGAC 120
QY	121	CAGTATGATTAAGCTACGAGGATTAATAATCGAAGAAGAACCAACGATGACACTTACC 180
Db	121	CAGTATGATTAAGCTACGAGGATTAATAATCGAAGAAGAACCAACGATGACACTTACC 180
QY	181	CATGCTGAGAATCCTGACTGGTACTACTCTGAGGCCAGCTGATCCATGCCAGCCCAAC 240
Db	181	CATGCTGAGAATCCTGACTGGTACTACTCTGAGGCCAGCTGATCCATGCCAGCCCAAC 240
QY	241	CCCTGTGAACACGGTGGGACTGCCTCGTCCATGGAGCACCTTACATGAGCTGCTG 300
Db	241	CCCTGTGAACACGGTGGGACTGCCTCGTCCATGGAGCACCTTACATGAGCTGCTG 300

1381 TCCGCGACGCTCCTGGATGCCAAAGTCAAGCTGATGCGAACACATTTGTGCAACTCCGCG 1440
Db TCCGCGACGCTCCTGGATGCCAAAGTCAAGCTGATGCGAACACATTTGTGCAACTCCGCG 1440
1441 CAACTCTATGACACATGATGATGACAGTATGATCTGTGAGGAAATCTTCAGAAACCT 1500
Db CAACTCTATGACACATGATGATGACAGTATGATCTGTGAGGAAATCTTCAGAAACCT 1500
1501 GGGCAAGACACCTGCCAGGCTGACTCTGTGAGGCCCCCTGACCTGTGAGAGGACGGAC 1560
Db GGGCAAGACACCTGCCAGGCTGACTCTGTGAGGCCCCCTGACCTGTGAGAGGACGGAC 1560
1561 TACTAGCTCTATGGATAGTGTGAGGCTGGGCTGGAGTGTGAGAGGACGGGCTCTAC 1620
Db TACTAGCTCTATGGATAGTGTGAGGCTGGGCTGGAGTGTGAGAGGACGGGCTCTAC 1620
1621 ACCCAAGTTACCAAAATTCCTGAATTTGGATCAAGGCCACCATCAAAAGTGAAGTGGCTTC 1680
Db ACCCAAGTTACCAAAATTCCTGAATTTGGATCAAGGCCACCATCAAAAGTGAAGTGGCTTC 1680
1681 TAA 1683
Db TAA 1683

RESULT 5
S83182
LOCUS 2408 bp mRNA linear PRI 11-FEB-1997
DEFINITION hyaluronan-binding protein-hepatocyte growth factor activator
ACCESSION S83182
VERSION S83182.1 GI:1836158
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1. (bases 1 to 2408)
Choi-Miura,N.H., Tobe,T., Sumiya,J., Nakano,Y., Sano,Y., Mazda,T.
and Tomita,M.
Purification and characterization of a novel hyaluronan-binding
protein (PHBP) from human plasma: it has three EGF, a kringle and a
serine protease domain, similar to hepatocyte growth factor
activator
J. Biochem. 119 (6), 1157-1165 (1996)
MEDLINE 96425001
PUBMED 8827452
REMARK GenBank staff at the National Library of Medicine created this
entry [NCBI Gibbsq 179630] from the original journal article.
This sequence comes from Fig. 3B.
Location/Qualifiers
1. .2408
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
1. .2408
/gene="hyaluronan-binding protein, PHBP"
97. .1779
/gene="hyaluronan-binding protein, PHBP"
/note="hepatocyte growth factor activator homolog; Method:
conceptual translation with partial peptide sequencing;
This sequence comes from Fig. 3B; PHBP"
/codon_start=1
/product="hyaluronan-binding protein"
/protein_id="AAB46909.1"
/db_xref="GI:1836158"
translation="MFARMSDLHVLLMALVGTAKTGFSLMSLLESIDPDWTFDQYDY
SYEDNGQENTSTLTAEHPDWTYTDQPCPNPCEHGGDCLVHGSFTFCSCILAP
FSGNKQVONTCKNDPCRGQCLITQSPYRCVKHPYTGSCQVVPVCRPNPCQ
NGATCSHRKSKFTCAQDFGKFCFCEIGSDDCYGDGYSYRGKNRTVNHACLYM
NSHLLQENTNMFEDAEATHGIEHNFRCNPDADKPFICFVTKNDKWEYCDVAC
SAQDVAYPESSPEPSKLPFGDSCGKTEIAERKIKRIYGGPKSTAGKHPQWASLOSS
LPLTISNPQGHFCGGALLHPCWLTAAHCTDIKTRHLKVLVGDQDLKKEEFHQSRV

EXIKVSHYNERDEIPHNDIALLLKLPVDGHCALESKYVKTCLPDGPPSPSSECHIS
GHWGYPETGKSGROLLDAKVLANTLNSRQLYDHWIDDSMICAGNLKPKGQDTCCGD
SGPLTCEKDGIIYVIGVSWGLECGKPGVTVTKPLNWLKIKESGF"

Query Match 99.8%; Score 1679.8; DB 9; Length 2408;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1681; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

ORIGIN

QY 1 ATGTTTCCAGGATGCTGATCTCCATGTTCTGCTGTTAATCGCTCTGGTGGGAAGACA 60
Db 97 ATGTTTCCAGGATGCTGATCTCCATGTTCTGCTGTTAATCGCTCTGGTGGGAAGACA 156
QY 61 GCCTTGGGTTCTCCCTGATGTTTATTGGAAAGCCTGGACCCAGATGGACCCCTGAC 120
Db 157 GCCTTGGGTTCTCCCTGATGTTTATTGGAAAGCCTGGACCCAGATGGACCCCTGAC 216
QY 121 CAGTATGATTACAGCTACGAGGATTTAATCAGAGAGAGAACACAGTAGACACACTTACC 180
Db 217 CAGTATGATTACAGCTACGAGGATTTAATCAGAGAGAGAACACAGTAGACACACTTACC 276
QY 181 CATGCTGAGAAATCCTGACTGGTACTACACTGAGGACCAAGCTGATCCATGCCAGCCCAAC 240
Db 277 CATGCTGAGAAATCCTGACTGGTACTACACTGAGGACCAAGCTGATCCATGCCAGCCCAAC 336
QY 241 CCTGTGAAACGGTGGGACTGCTGCTCCATGGGAGCAGCTTCACATGACAGTCCCTG 300
Db 337 CCTGTGAAACGGTGGGACTGCTGCTCCATGGGAGCAGCTTCACATGACAGTCCCTG 396
QY 301 GCTCTTTTCTCTGGGAATAAGTGTGAGAAAGTGCAGAAATACGTCGAAGGACCAACCCATGT 360
Db 397 GCTCTTTTCTCTGGGAATAAGTGTGAGAAAGTGCAGAAATACGTCGAAGGACCAACCCATGT 456
QY 361 GGCGGGGCCAATGTCATTACCCAGAGTCTCCCTACTACCGCTGTGTCTTAACAC 420
Db 457 GGCGGGGCCAATGTCATTACCCAGAGTCTCCCTACTACCGCTGTGTCTTAACAC 516
QY 421 CCTTACACAGTCCAGCTGCTCCCAAGTGGTTCCTGTATGAGGACCAACCCCTGCCAG 480
Db 517 CCTTACACAGTCCAGCTGCTCCCAAGTGGTTCCTGTATGAGGACCAACCCCTGCCAG 576
QY 481 AATGGGCTACCTGCTCCCGCATAAAGCGAGATCCAAAGTTCACCTGTGCTGCTCCGAC 540
Db 577 AATGGGCTACCTGCTCCCGCATAAAGCGAGATCCAAAGTTCACCTGTGCTGCTCCGAC 636
QY 541 CAGTTCAGGGGAAATTCGTGAATAGTTCGTGATGACTGCTATGTTGGCGATGGCTAC 600
Db 637 CAGTTCAGGGGAAATTCGTGAATAGTTCGTGATGACTGCTATGTTGGCGATGGCTAC 696
QY 601 TCTTACCGAGGAAATGAATAGACAGTCAACAGCATGCGTCCCTTTACTGGAATCC 660
Db 697 TCTTACCGAGGAAATGAATAGACAGTCAACAGCATGCGTCCCTTTACTGGAATCC 756
QY 661 CACCTCTCTTGCAGGAGAAATTAACAATGTTTATGAGGATGCTGAAACCCATGGGATT 720
Db 757 CACCTCTCTTGCAGGAGAAATTAACAATGTTTATGAGGATGCTGAAACCCATGGGATT 816
QY 721 GGGGAACACAAATTTCTGCAGAAACCCAGATGCGGACGAAAGCCCTGGTGTATTATAA 780
Db 817 GGGGAACACAAATTTCTGCAGAAACCCAGATGCGGACGAAAGCCCTGGTGTATTATAA 876
QY 781 GTTACCAATGACAGGTGAATGGAATACGTGATCTCTCAGCTCTCAGCCAGCCAGAC 840
Db 877 GTTACCAATGACAGGTGAATGGAATACGTGATCTCTCAGCTCTCAGCCAGCCAGAC 936
QY 841 GTTGCTTACCCAGGAAAGCCCACTGAGCCATCAACCAAGCTTCGCGGGTTTGACTCC 900
Db 937 GTTGCTTACCCAGGAAAGCCCACTGAGCCATCAACCAAGCTTCGCGGGTTTGACTCC 996
QY 901 TGTGAAAGACTGAGATAGCAGAGAGAGAGATCAAGAGAAATCTATGGAGGCTTTAAGAGC 960
Db 997 TGTGAAAGACTGAGATAGCAGAGAGAGAGATCAAGAGAAATCTATGGAGGCTTTAAGAGC 1056

961 ACGGCGGGCAAGCACCATGGCAGGGCTCCCTCCAGCTCTCGCTCGCTCTGACCATCTCC 1020
Db
1057 ACGGCGGGCAAGCACCATGGCAGGGCTCCCTCCAGCTCTCGCTCGCTCTGACCATCTCC 1116
Qy 1021 ATGCCCCAGGGCCACCTTCTGTGGTGGGGCCCTGATCCACCCCTGCTGGGTGCTCATGCT 1080
Db 1117 ATGCCCCAGGGCCACCTTCTGTGGTGGGGCCCTGATCCACCCCTGCTGGGTGCTCATGCT 1176
Qy 1081 GCCCACTGCACCCGACATAAAAAACAGACATCTAAAGGTGGTGTAGGGACACAGGACCTG 1140
Db 1177 GCCCACTGCACCCGACATAAAAAACAGACATCTAAAGGTGGTGTAGGGACACAGGACCTG 1236
Qy 1141 AAGAAAGAGAAATTCATGAGCAGAGCTTTAGGGTGCAGAGATATTCAGATGACGCCAC 1200
Db 1237 AAGAAAGAGAAATTCATGAGCAGAGCTTTAGGGTGCAGAGATATTCAGATGACGCCAC 1296
Qy 1201 TACAATCAAGAGATGAGATTCCCAACATGATATTCATGCTCAAGTTAAAGCCAGTG 1260
Db 1297 TACAATCAAGAGATGAGATTCCCAACATGATATTCATGCTCAAGTTAAAGCCAGTG 1356
Qy 1261 GATGGTCACTGTGCTTAGAATCAAAATACGTGAAGACTGTGCTGCTGATGGGTCC 1320
Db 1357 GATGGTCACTGTGCTTAGAATCAAAATACGTGAAGACTGTGCTGCTGATGGGTCC 1416
Qy 1321 TTTCCCTCTGGAGTGAGTGCACATCTCTGGCTGGGGTGTACAGAAACAGGAAAGGG 1380
Db 1417 TTTCCCTCTGGAGTGAGTGCACATCTCTGGCTGGGGTGTACAGAAACAGGAAAGGG 1476
Qy 1381 TCCGCCAGCTCTCGGATGCCAAAGTCAAGTGATTGCCAACACTTTGTCAACTCCCGC 1440
Db 1477 TCCGCCAGCTCTCGGATGCCAAAGTCAAGTGATTGCCAACACTTTGTCAACTCCCGC 1536
Qy 1441 CAACTCTATGACCATGATGATGACAGATGATCTGTCGCGGAATCTTCAGAAACCT 1500
Db 1537 CAACTCTATGACCATGATGATGACAGATGATCTGTCGCGGAATCTTCAGAAACCT 1596
Qy 1501 GGGCAAGACACCTGCCAGGGTGACTCTGGAGGCCCTGACCTGTGAGAGGACGGCACC 1560
Db 1597 GGGCAAGACACCTGCCAGGGTGACTCTGGAGGCCCTGACCTGTGAGAGGACGGCACC 1656
Qy 1561 TACTACGTCTATGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1620
Db 1657 TACTACGTCTATGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1716
Qy 1621 ACCCAAGTTACAAATTCCTGAATTCGATCAAGCCACCATCAAAAGTGAAGTGGCTTC 1680
Db 1717 ACCCAAGTTACAAATTCCTGAATTCGATCAAGCCACCATCAAAAGTGAAGTGGCTTC 1776
Qy 1681 TAA 1683
Db 1777 TAA 1779

RESULT 6
AX409021
LOCUS AX409021 3008 bp DNA linear PAT 14-JUN-2002
DEFINITION Sequence 1668 from Patent WO0229103.
ACCESSION AX409021
VERSION AX409021.1 GI:21441726
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1
REFERENCE
AUTHORS Alves, C., Horne, D., Peres-da-Silva, S. and Vockley, J.G.
TITLE Gene expression profiles in liver cancer
JOURNAL Patent: WO 0229103-A 1668 11-APR-2002;
GENE LOGIC INC (US)
FEATURES
Location/Qualifiers
1..3008
/organism="Homo sapiens"
/mol_type="unassigned DNA"

/db_xref="taxon:9606"
/note="EMBL/GenBank Accession No. D49742"
Query Match 99.8%; Score 1679.8; DB 6; Length 3008;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1681; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 1 ATGTTTGCAGGATGTCTGATCTCCATGTTCTGCTGTTTAAATGGTCTGGTGGGAAAGACA 60
Db 97 ATGTTTGCAGGATGTCTGATCTCCATGTTCTGCTGTTTAAATGGTCTGGTGGGAAAGACA 156
Qy 61 GCCTGTGGGTTCCTCCCTGATCTCTTATTTGAAAGCCTGGACCCAGACTGACCCCTGAC 120
Db 157 GCCTGTGGGTTCCTCCCTGATCTCTTATTTGAAAGCCTGGACCCAGACTGACCCCTGAC 216
Qy 121 CAGTATGATTACAGCTACGAGGATTTAATCAGGAAGAGAAACACAGTAGCACATTACC 180
Db 217 CAGTATGATTACAGCTACGAGGATTTAATCAGGAAGAGAAACACAGTAGCACATTACC 276
Qy 181 CATGTGAGATTCCTGACTGTGTACTACACTGAGACCAAGCTGATCCATGCCAGCCCAAC 240
Db 277 CATGTGAGATTCCTGACTGTGTACTACACTGAGACCAAGCTGATCCATGCCAGCCCAAC 336
Qy 241 CCTGTGAAACCGTGGGGACTGCTCCTCCATGGGAGCACCTTCACATGCAAGTGCCTG 300
Db 337 CCTGTGAAACCGTGGGGACTGCTCCTCCATGGGAGCACCTTCACATGCAAGTGCCTG 396
Qy 301 GCTCTTTCTCTGGAATTAAGTGTGAGAAAGTGCAAAATAGTGTGCAAGGACAAACCTATG 360
Db 397 GCTCTTTCTCTGGAATTAAGTGTGAGAAAGTGCAAAATAGTGTGCAAGGACAAACCTATG 456
Qy 361 GSCCGGGGCCAATGCTCATTTACCCAGAGTCTCCTCCTACTACCGCTGTGTGTAAACAC 420
Db 457 GSCCGGGGCCAATGCTCATTTACCCAGAGTCTCCTCCTACTACCGCTGTGTGTAAACAC 516
Qy 421 CTTTACACAGTCTCCAGTGTCTCCAAAGTGGTTCCTGTATGCAAGGCCAAAACCTTGCAG 480
Db 517 CTTTACACAGTCTCCAGTGTCTCCAAAGTGGTTCCTGTATGCAAGGCCAAAACCTTGCAG 576
Qy 481 AATGGGCTACCTGCTCCCGCATAGCGGAGATCCAAAGTTCACTGTGCTGTCCCTGAC 540
Db 577 AATGGGCTACCTGCTCCCGCATAGCGGAGATCCAAAGTTCACTGTGCTGTCCCTGAC 636
Qy 541 CAGTTCAAGGGGAAATTTCTGAAATAGTGTGTGATGATGCTATGTTGGGATGGCTAC 600
Db 637 CAGTTCAAGGGGAAATTTCTGAAATAGTGTGTGATGATGCTATGTTGGGATGGCTAC 696
Qy 601 TCTTACCGAGGAAATGATAGGACAGTCAACAGCATGGTGGCTTTTACTGGAACTCC 660
Db 697 TCTTACCGAGGAAATGATAGGACAGTCAACAGCATGGTGGCTTTTACTGGAACTCC 756
Qy 661 CACCTCCTCTTGCAGGAAATTAACCATGTTTATGGAGGATGCTGAAACCCATGGGATT 720
Db 757 CACCTCCTCTTGCAGGAAATTAACCATGTTTATGGAGGATGCTGAAACCCATGGGATT 816
Qy 721 GGGGAAACAATTTCTGAGAAACCCAGATGCGGAGGAAAGCCCTGTGCTTTTAA 780
Db 817 GGGGAAACAATTTCTGAGAAACCCAGATGCGGAGGAAAGCCCTGTGCTTTTAA 876
Qy 781 GTTACCAATGACAAAGGTGAAATGGGAATACTGCTGATGCTCAGCTGTCTCAGCCAGGAC 840
Db 877 GTTACCAATGACAAAGGTGAAATGGGAATACTGCTGATGCTCAGCTGTCTCAGCCAGGAC 936
Qy 841 GTTGTCTTACCCAGAGGAAAGCCCACTGAGCCATCAACCAAGCTTCGCGGGTTTACTCC 900
Db 937 GTTGTCTTACCCAGAGGAAAGCCCACTGAGCCATCAACCAAGCTTCGCGGGTTTACTCC 996
Qy 901 TGTGGAAGACTGAGATGACAGAGGAGAGATCAAGAGATCTATGAGGCTTTTAAGAGC 960
Db 997 TGTGGAAGACTGAGATGACAGAGGAGAGATCAAGAGATCTATGAGGCTTTTAAGAGC 1056
Qy 961 ACGGCGGGCAAGCACCCATGGCAGGGCTCCCTCCAGTCTCGCTCTGACCATCTCC 1020

Db	1057	ACGGCGGCAAGACCCATGAGCGCTCCCTCCAGTCTCGCTGCTGACCATCTCC	1116
Qy	1021	ATGCCCCAGGCGACCTTCTGTGGTGGGCGCTGATCCACCCCTGCTGGTGTCTACTGCT	1080
Db	1117	ATGCCCCAGGCGACCTTCTGTGGTGGGCGCTGATCCACCCCTGCTGGTGTCTACTGCT	1176
Qy	1081	GCCCACTGACCGACATATAAAACCCAGACATCTAAAGGTGGTCTAGGCGACCCAGGACCTG	1140
Db	1177	GCCCACTGACCGACATATAAAACCCAGACATCTAAAGGTGGTCTAGGCGACCCAGGACCTG	1236
Qy	1141	AGAAAGAGAAATTCATGAGCAGAGCTTTAGGGTGCAGAGATATTCAGTACAGCCAC	1200
Db	1237	AGAAAGAGAAATTCATGAGCAGAGCTTTAGGGTGCAGAGATATTCAGTACAGCCAC	1296
Qy	1201	TACAATGAAGAGATGAGATTCACCAATGATATTCGATTCCTCAAGTAAAGCCAGCTG	1260
Db	1297	TACAATGAAGAGATGAGATTCACCAATGATATTCGATTCCTCAAGTAAAGCCAGCTG	1356
Qy	1261	GATGTCATCTGCTCTAGATCCAAATACGTGAAGACATGTGCTTGCCTGATGGGTCC	1320
Db	1357	GATGTCATCTGCTCTAGATCCAAATACGTGAAGACATGTGCTTGCCTGATGGGTCC	1416
Qy	1321	TTTCCCTCTGGAGTGGAGTGCACATCTCTGGCTGGGGTGTACAGAAACAGGAAAGGG	1380
Db	1417	TTTCCCTCTGGAGTGGAGTGCACATCTCTGGCTGGGGTGTACAGAAACAGGAAAGGG	1476
Qy	1381	TCCCGCAGCTCTGATGATCCAAAGTCAAGCTGATGTCACACATTTGTGCAACTCCGC	1440
Db	1477	TCCCGCAGCTCTGATGATCCAAAGTCAAGCTGATGTCACACATTTGTGCAACTCCGC	1536
Qy	1441	CAACTCTATGACCATGATGATGACAGTATGATCTGTGACAGGAAATCTTCAGAAACCT	1500
Db	1537	CAACTCTATGACCATGATGATGACAGTATGATCTGTGACAGGAAATCTTCAGAAACCT	1596
Qy	1501	GGGCAAGACACCTGCGAGGTGACTCTGGAGCCGCCCTGACCTGTGAGAGAGCGGCACC	1560
Db	1597	GGGCAAGACACCTGCGAGGTGACTCTGGAGCCGCCCTGACCTGTGAGAGAGCGGCACC	1656
Qy	1561	TACTAGCTCTATGGGATAGTACGCTGGGCGCTGGAGTGTGAGAGAGCGGCTCTAC	1620
Db	1657	TACTAGCTCTATGGGATAGTACGCTGGGCGCTGGAGTGTGAGAGAGCGGCTCTAC	1716
Qy	1621	ACCAAGTTACCAAAATTCCTGAATTGGATCAAGGCCACCATCAAAAGTGAAGTGGCTTC	1680
Db	1717	ACCAAGTTACCAAAATTCCTGAATTGGATCAAGGCCACCATCAAAAGTGAAGTGGCTTC	1776
Qy	1681	TAA 1683	
Db	1777	TAA 1779	

RESULT 7
HUMHGFAL
LOCUS
DEFINITION
HUMAN mRNA for HGF activator like protein, complete cds.
ACCESSION
D49742
VERSION
D49742.1
KEYWORDS
HGF activator like protein; serin protease.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 3008)
Kitamura,N.
Unpublished
REFERENCE
2 (bases 1 to 3008)
Kitamura,N.
Direct Submission
TITLE
Submitted (17-MAR-1995) Naomi Kitamura, Institute for Liver
Research, Kansai Medical University; Moriguchi, Osaka 570, Japan
(Tel:06-992-1001(ex.2530), Fax:06-994-6099)
Location/Qualifiers

source	1.	3008	/organism="Homo sapiens"	
			/mol_type="mRNA"	
			/db_xref="taxon:9606"	
CDS		97..1779	/codon_start=1	
			/product="HGF activator like protein"	
			/protein_id="BAA08576.1"	
			/db_xref="GI:1345398"	
			/translation="MFARVSDLHVLMLVGTACGFSLSLLSLSDPDTPDOYDY SVEDYNQENTSTLTHAENPDWYTEDOAPCPNPEHGGDCLVHGSTFTCSCLAP FSGKCKQVNTCKNPNRQCGCLIQSPFYRCVCKHYTPGFSQVQVPCRPQCY NGATSRHKRSKFTCCPDQKGFCEGSDCVVGGDSYRGNQNVTVNGHACLYQ NSHLLQENNMFMEDAETHGICENFNDADEKPCFVKITNDKXVGECDVSAC SADVAYPESTPESTKLPDFDSCGCTIAERKIKRIFGFGKSTAGKGPWASQSS LPTISMPQCHFCGAGALHPWLVTAHCTDIKTRHLKVLGDLQDLKKEEFHEQSPRV LETKFISHYNERDEI PHNDIALLKCPVDGHCHALSKYKVTCLPDGSPSPSECHIS GNGVETGKSGRLLDKVKLIJANTLCNSRQLYDHMDIMICAGNLQKPGDTCQGD SGPLICEKDKGYVYVYVSVGLECGKRPVGTQVTKFLNWKATIKSESSEF"	
		2312	polyA_site	
		3008	polyA_site	
ORIGIN				
			Query Match	99.8%; Score 1679.8; DB 9; Length 3008;
			Best Local Similarity	99.9%; Pred. No. 0;
			Matches 1681; Conservative	0; Mismatches 2; Indels 0; Gaps 0;
Qy	1	ATGTTTGCCAGGATGCTGATCTCCATGTTCTGCTCTTAAATGGCTCTGTTGGGAAAGACA	60	
Db	97	ATGTTTGCCAGGATGCTGATCTCCATGTTCTGCTCTTAAATGGCTCTGTTGGGAAAGACA	156	
Qy	61	GCCTGTGGGTTCCTCCCTGATGCTTTATTGGAAAGCTGACCCAGACTGGACCCCTGAC	120	
Db	157	GCCTGTGGGTTCCTCCCTGATGCTTTATTGGAAAGCTGACCCAGACTGGACCCCTGAC	216	
Qy	121	CAGTATGATTACAGCTACGAGGATTAATACAGAAAGAGAACACCACTAGTACACTTACC	180	
Db	217	CAGTATGATTACAGCTACGAGGATTAATACAGAAAGAGAACACCACTAGTACACTTACC	276	
Qy	181	CATGCTGAGAAATCCTGACTGCTACTACCTGAGAGCAAGCTGATCCATGCGAGCCCAAC	240	
Db	277	CATGCTGAGAAATCCTGACTGCTACTACCTGAGAGCAAGCTGATCCATGCGAGCCCAAC	336	
Qy	241	CCCTGTGAACACAGGTGGGACTCCCTCGTCCATGGAGCACTTACATGAGCTGCTG	300	
Db	337	CCCTGTGAACACAGGTGGGACTCCCTCGTCCATGGAGCACTTACATGAGCTGCTG	396	
Qy	301	GCTCTTTCTCTGGGATAGTCTCAGAAAGTGCRAAATACGTGCAAGGACCAACCCATGT	360	
Db	397	GCTCTTTCTCTGGGATAGTCTCAGAAAGTGCRAAATACGTGCAAGGACCAACCCATGT	456	
Qy	361	GGCGGGGCCCAATGCTCTCATTACCCAGAGTCTTCCCTACTACCGCTGTGTCTGTAACAC	420	
Db	457	GGCGGGGCCCAATGCTCTCATTACCCAGAGTCTTCCCTACTACCGCTGTGTCTGTAACAC	516	
Qy	421	CTTTACAGAGTCCAGCTGCTCCCAAGTGTTCCTGTATGAGGCGCAACCCCTGCCAG	480	
Db	517	CTTTACAGAGTCCAGCTGCTCCCAAGTGTTCCTGTATGAGGCGCAACCCCTGCCAG	576	
Qy	481	AATGGGCTACCTGCTCCCGGCATAGCGGAGATCCAAGTTCACTGTGCTGCTGCCGAC	540	
Db	577	AATGGGCTACCTGCTCCCGGCATAGCGGAGATCCAAGTTCACTGTGCTGCTGCCGAC	636	
Qy	541	CAGTTCAAGGGGAAATTCCTGGAATAGGTTCTGTATGAGTGTCTGTTGGCGATGCTAC	600	
Db	637	CAGTTCAAGGGGAAATTCCTGGAATAGGTTCTGTATGAGTGTCTGTTGGCGATGCTAC	696	
Qy	601	TCCTTACCGAGGAAATGAATAGACAGTCAACAGCATCGCTGCCTTTACTGGAATCC	660	
Db	697	TCCTTACCGAGGAAATGAATAGACAGTCAACAGCATCGCTGCCTTTACTGGAATCC	756	
Qy	661	CACCTCTCTTGCAGGAGAAATTAACAATCTTTATGAGGATGCTGAAACCCATGGGAT	720	

Db 757 CACCTCTCTTGGAGGAGATTTACAACATGTTTATGAGGAGTGTGTAACCCATGGGATT 816
Qy 721 GGGGACACAAATTTCTGAGAAACCCAGATCGGACGAAAGCCCTGCTTATTTAA 780
Db 817 GGGGACACAAATTTCTGAGAAACCCAGATCGGACGAAAGCCCTGCTTATTTAA 876
Qy 781 GTTACCAATGACAAGGTGAATGGAATCTGTGATGTCTCAGCCTGCTCAGCCAGGAC 840
Db 877 GTTACCAATGACAAGGTGAATGGAATCTGTGATGTCTCAGCCTGCTCAGCCAGGAC 936
Qy 841 GTTGCTACCCAGAGAAAGCCCTGAGCCATCAACCAAGCTCCGGGTTGACTCC 900
Db 937 GTTGCTACCCAGAGAAAGCCCTGAGCCATCAACCAAGCTCCGGGTTGACTCC 996
Qy 901 TGTGAAAGACTGAGATACAGAGAGAGATCAAGAGAAATCTATGGAGGCTTTAAGAGC 960
Db 997 TGTGAAAGACTGAGATACAGAGAGAGATCAAGAGAAATCTATGGAGGCTTTAAGAGC 1056
Qy 961 AGGGGGGGAAGCAACCCATGAGAGGCTCCCTCAGTCTCGCTGCTGCTGACTCTCC 1020
Db 1057 AGGGGGGGAAGCAACCCATGAGAGGCTCCCTCAGTCTCGCTGCTGCTGACTCTCC 1116
Qy 1021 ATGCCCCAGGGCCACTTCTGTGGTGGGGGCTGATCCACCCTGCTGGTGTCTCACTGCT 1080
Db 1117 ATGCCCCAGGGCCACTTCTGTGGTGGGGGCTGATCCACCCTGCTGGTGTCTCACTGCT 1176
Qy 1081 GGCCTCTGACCAACCAATGAGATTTCCCAATGATATTTGCAATCTCAAGTTAAAGCCAGTG 1140
Db 1177 GGCCTCTGACCAACCAATGAGATTTCCCAATGATATTTGCAATCTCAAGTTAAAGCCAGTG 1236
Qy 1141 AAGAAGAGAGATTTTCATGAGAGAGCTTTAGGGTGCAGAGATATTTCAAGTACAGCCAC 1200
Db 1237 AAGAAGAGAGATTTTCATGAGAGAGCTTTAGGGTGCAGAGATATTTCAAGTACAGCCAC 1296
Qy 1201 TACAATGAAAGAGATGAGATTTCCCAATGATATTTGCAATCTCAAGTTAAAGCCAGTG 1260
Db 1297 TACAATGAAAGAGATGAGATTTCCCAATGATATTTGCAATCTCAAGTTAAAGCCAGTG 1356
Qy 1261 GATGTCATGCTCTAGATCCCAATGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1320
Db 1357 GATGTCATGCTCTAGATCCCAATGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1416
Qy 1321 TTTCCCTCTGGGAGTGAAGTGCACATCTCTGGCTGGGGTGTACAGAAACAGGAAAGGG 1380
Db 1417 TTTCCCTCTGGGAGTGAAGTGCACATCTCTGGCTGGGGTGTACAGAAACAGGAAAGGG 1476
Qy 1381 TCCCGCCAGCTCTGATGCGCAAGTCAAGCTGATGCGCAACACATTTGCGCACTCCGCG 1440
Db 1477 TCCCGCCAGCTCTGATGCGCAAGTCAAGCTGATGCGCAACACATTTGCGCACTCCGCG 1536
Qy 1441 CAACTCTATGACCAATGATTTGATGACAGTATGATCTGTGCGAGGAAATCTTCAGAAACCT 1500
Db 1537 CAACTCTATGACCAATGATTTGATGACAGTATGATCTGTGCGAGGAAATCTTCAGAAACCT 1596
Qy 1501 GGGCAAGACACTGCGCAGGGTGAATCTGAGGCCCCCTGACCTGTGAGAGGAGCGGAC 1560
Db 1597 GGGCAAGACACTGCGCAGGGTGAATCTGAGGCCCCCTGACCTGTGAGAGGAGCGGAC 1656
Qy 1561 TACTAGCTCTATGGGATGAGCTGGGCTGGAGTGTGAGAGAGGCGGAGGCTCTAC 1620
Db 1657 TACTAGCTCTATGGGATGAGCTGGGCTGGAGTGTGAGAGAGGCGGAGGCTCTAC 1716
Qy 1621 ACCCAAGTACCAAAATTCCTGAATGGATCAAGCCCATCAAAAGTGAAGAGTGCTTC 1680
Db 1717 ACCCAAGTACCAAAATTCCTGAATGGATCAAGCCCATCAAAAGTGAAGAGTGCTTC 1776
Qy 1681 TAA 1683
Db 1777 TAA 1779

RESULT 8
BC031412

LOCUS BC031412 2251 bp mRNA linear PRI 07-OCT-2003
DEFINITION Homo sapiens hyaluronan binding protein 2, mRNA (cDNA clone MGC:34485 IMAGE:5187066), complete cds.
ACCESSION BC031412
VERSION BC031412.1 GI:21618648
KEYWORDS MGC.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
MAMMALIA; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 2251)
Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Klausner R.D., Collins F.S., Wagner J., Shenmen C.M., Schuler G.D., Altshuler S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K., Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F., Datchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L., Sapotnick L., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E., Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C., Raha S.S., Loquellano N.A., Peters G.J., Aronson R.D., Mullany S.J., Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H., Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W., Viallon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A., Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A., Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G., Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C., Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S., Krzywinski M.I., Skalska U., Smalios D.E., Schnurch A., Schein J.E., Jones S.J. and Marra M.A.
TITLE Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
MEDLINE 22388257
PUBMED 12477932
REFERENCE 2 (bases 1 to 2251)
Strausberg R.
Direct Submission
Submitted (06-JUN-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>
COMMENT Contact: MGC help desk
Email: cgabs-rc@mail.nih.gov
Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA sequencing by: Baylor College of Medicine Human Genome Sequencing Center
Center code: BCM-HGSC
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>
Contact: amg@bcm.tmc.edu
Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Lounsged, H., Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati, A.N., Gibbs, R.A.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAK Plate: 50 Row: g Column: 21
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 20302151.
FEATURES
Location/Qualifiers
1..2251
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="MGC:34485 IMAGE:5187066"
/tissue_type="Colon, Kidney, Stomach, adult, whole pooled"
/clone_id="NIH_MGC_116"
/lab_host="DH10B"
/note="vector: pCMV-SPORT6"
1..2251
/gene="HABP2"

[illegible]

```
RESULT 9
AKI28915
LOCUS
DEFINITION
Mus musculus cDNA fis, clone TRACH2024735, moderately similar to
Homo sapiens hyaluronan binding protein 2 (HABP2).
ACCESSION
AKI28915
VERSION
AKI28915.1 GI:34536572
KEYWORDS
oligo capping; fis (full insert sequence).
SOURCE
Mus musculus (house mouse)
ORGANISM
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1
Ninomiya, K., Wagatsuma, M., Kanda, K., Kondo, H., Yokoi, T.,
Kodaira, H., Furuya, T., Takahashi, M., Kikkawa, E., Omura, Y., Abe, K.,
Kamihara, K., Katsuta, N., Sato, K., Tanikawa, M., Yamazaki, M.,
Sugiyama, T., Irie, R., Otsuki, T., Sato, H., Ota, T., Wakamatsu, A.,
Ishii, S., Yamamoto, J., Isono, Y., Kawai-Hio, Y., Saito, K.,
Nishikawa, T., Kimura, K., Yamashita, H., Matsuo, K., Nakamura, Y.,
Sekine, M., Kikuchi, H., Murakawa, K., Kanehori, K.,
Takahashi-Fujii, A., Oshima, A., Sugiyama, A., Kawakami, B., Suzuki, Y.,
Sugano, S., Nagahara, K., Masuho, Y., Nagai, K. and Isogai, T.
NEDO cDNA sequencing project
Unpublished
2 (bases 1 to 2190)
Isogai, T. and Yamamoto, J.
Direct Submission
Submitted (15-JUL-2003) Takao Isogai, FLJ Project (HRI Team); 2-6-7
Kazusa-Kamatari, Kisarazu, Chiba 292-0818, Japan
(E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)
NEDO cDNA sequencing project supported by Ministry of Economy,
Trade and Industry of Japan; cDNA full insert sequencing; Research
Association for Biotechnology (RAB); cDNA library construction;
Helix Research Institute (HRI) (supported by Japan Key Technology
Center etc.); 5'- & 3'-end one pass sequencing; RAB, HRI, and
Biotechnology Center, National Institute of Technology and
Evaluation; Clone selection for full insert sequencing: HRI and
RAB; annotation: HRI and RAB.
Location/Qualifiers
1..2190
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="TRACH2024735"
/note="cloning vector: pME18SFL3"

ORIGIN
Query Match 67.8%; Score 1141; DB 10; Length 2190;
Best Local Similarity 80.3%; Pred. No. 0;
Matches 1352; Conservative 0; Mismatches 325; Indels 6; Gaps 1;

QY 1 ATGTTTGCAGGATGCTGATCTCCATGTTCTGCTGTTAATGGCTCTGGTGGGAAGACA 60
DB 62 ATATTTGTCCAGATGTTGGTGTTCGTGCTGCTGCTAATCGCCCTGGTGGGAAGTCA 121
QY 61 GCTGTGGGTTCTCCCTGATGCTTTATTGGAAGCCTGGACCCAGACTGGACCCCTGAC 120
DB 122 GTCAATTGGGCTCTCACTGATGCTTTCAATGCGCCCCAGACCCAGATTGGACCCCGAT 181
QY 121 CAGTATGATTACAGCTACGAGGATTATATACAGGAAGAACACCAAGTAGCACCTTACC 180
DB 182 GACTATTACTACAGCTATGAGCAGTCCAGCCAGACGAAGACCCAGTGTCAAGCAGACC 241
QY 181 CATGCTGAGATCTGACTGGTACTACACTGAGGACCAAGCTGATCCATGCCAGCCCAAC 240
DB 242 ACCCTTGAGAACCCGACCTGGTACT-----ATGAAGACGATGATCCATGCCAGTCCAAC 295
QY 241 CCCTGTGAACACGGTGGGACTGCCTCGTCCATGGGAGCACCTTCACATGCAAGCTGCCTG 300
DB 296 CCCTGTGAACACGGGAGTGTATCATCAGGGGATACCTTCAGTTGCAAGCTGCCCA 355
```

```
QY 301 GCTCCTTCTCTGGGATAAGTGTGAGAAAGTGCAGAAATACGTGCAAGACCAACCATGT 360
DB 356 GCCCCCTTCTCGGGAGCCGGTGCCAGACTGCACAAAAACAAGTGAAGACAAACCCATGT 415
QY 361 GGCGGGGGCAATGTCTCATTAACCCAGAGTCTCCTACTACCGCTGTGTCTGTAAACAC 420
DB 416 GTCATGGTGAATTGCTCTCATTAACCCAGAGACCCCTACTACCGCTGTGTGCAATAC 475
QY 421 CTTTACACAGGTCCAGCTGCTCCCAAGTGGTTCCTGTATGAGGACCAACCCCTGCCAG 480
DB 476 CTTTACACGGGACCAAGCTGCTCCAAAGTGTTCGGGCATGAGGCAACCCCTGCCAG 535
QY 481 AATGGGGCTACCTGCTCCCGGCATAGCGGAGATCCAAGTTCACCTGTGCTGTCTCCGAC 540
DB 536 AATGGGGAGTCTGTTCCCGACACAGCGAGATCCAGGTTTACCTGTGCTGTCTCCAGAC 595
QY 541 CAGTCAAGGGGAAATCTGTGAATAGTGTCTGTAGTACTGTATGTTGGCGATGCTTAC 600
DB 596 CAGTAAAGGGGAAATCTGTGAATAGTGTCCGAGACTGTGTATGTCGGTGTATGCTTAC 655
QY 601 TCTTACCGAGGGAAATGAATAGGACAGTCAACCAAGCATGCGTGCCTTTACTTGGAACTCC 660
DB 656 TCTTACCGAGGCAAGTGAATAGACAGTCAACCAAGCATGCGTGCCTTTACTTGGAACTCC 715
QY 661 CACTCTCTTGTGAGGAGAAATTACAAATGTTTATGAGGATGCTGAAACCCATGGGAT 720
DB 716 CACTCTCTTGTGAGGAGACTTATAACATGTTTATGAGGATGCGAGACCCACCGGATC 775
QY 721 GGGGAACACAAATTTCTGCAGAAACCCAGATGCGAGCAAGCCCTGGTGTCTTATATAA 780
DB 776 GCAGAGCAACAATCTCTGCAGAAACCCAGATGAGGACCAACAACCTGGTGTCTCGTCAAG 835
QY 781 GTTACCAATGACAAGGTGAATAGGAAATCTGTGATGTCTAGCTGTCTCAGCCACAGGAC 840
DB 836 GTGAACAGTGAAGGTGAATAGGAAATCTGTGATGTCTACAGTCTGTCCAGTGCCTGAC 895
QY 841 GTTGCCTTACCAGAGAAAGCCCACTGAGGACATCAACCAAGCTTCCGGGTTTGACTCC 900
DB 896 ACCCTTAACCAAGTGAAGGCTTCTGAGGCTGTGATGAGCTGCAGGTTCCAGTCTC 955
QY 901 TGTGGAAGAAGCTGAGATAGCAGAGAGGAAGATCAAGAGAATCTATGAGGCTTTAAGAGC 960
DB 956 TGCAGGAAGACGAGGTAGTGAACACGAGTCAAGCGTATCTACGGGGCTTTAAGAGC 1015
QY 961 ACGGGGGAGACCAACCATGAGGAGGCTGCTCCAGTCTCGTCTGCTGCTGACCTCTCC 1020
DB 1016 ACAGCAGGCAAGCACCCGTCGAGGTGTCCTCTGAGACCTCACTGCGGTGTACCACTCC 1075
QY 1021 ATGCCCCAGGCGCACTTCTGTGTGGGGGCTGATCCACCCCTGCTGGGTGCTCACTGCT 1080
DB 1076 ATGCCCCAGGCGCACTTCTGTGGGGGCGCCCTGATCCACCCCTGCTGGGTGCTCACTGCA 1135
QY 1081 GCGCCTGCAACGACATAAAACACAGACATCTAAAGTGGTGTCTAGGGACCAAGGACCTG 1140
DB 1136 GCGCCTGTACCGACATAAACACCAAGCATCTAAAGTGTGTACTAGGGGATCAGGACCTG 1195
QY 1141 AAGAAAGAAGAAATTTATGAGCAGAGCTTTAGGGTGCAGAAATATTTCAAGTACAGCCAC 1200
DB 1196 AAGAGACAGAAATCCCATGACAGACCTTCAGGGTGGAAAAATACTGAATACAGTACAG 1255
QY 1201 TACAATGAAGAGATGAGATTCCTCCCAATGATATTGCTAATGCTAAGTTAAAGCCAGTG 1260
DB 1256 TATAATGAAGAGATGAGATTCCTCCCAATGATATTGCTTGTCAAGTTAAAGCCAGTG 1315
QY 1261 GATGGTCACTGTCTCTAGAAATCCAATACGTGAAGACTGTGTGCTTGTGCTGATGGTCC 1320
DB 1316 GTGGTCACTGTCTCTGGAAATCCAGATATGTAAGACTGTATGTTGTCAGGACCC 1375
QY 1321 TTTCCCTCTCGGAGTGAAGTCCACATCTCTGGCTGGGGTGTATACAGAAACAGAAAGGG 1380
DB 1376 TTTCCCTCTGAACTGAGTGCCACATCTCTGGCTGGGGTGTATACAGAAACAGGGAAGGG 1435
QY 1381 TCCGCCAGCTCTCTGGATGCCAAGTCAAGCTGATGCCAACACTTGTGCAACTCCCGC 1440
```

CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)

1541 CCTGTGAGAGACGGCCTACTACTCTATGGATGATGAGCTGGGCTGGAGTGTG 1600
 1470 CCTGTGAGAGATGGAATCTACTACTCTAGCTACGGGATTTGTAAGCTGGGCGCCAGGAATGTG 1529
 1601 AGAAGAGGCGGAGGTCTACACCCCAAGTTACCAAAATCTCTGAATTCGATCAAGACCA 1660
 1530 GGAAGAGCGGAGGTCTACACTCAAGTCACCAAGTCTCTGAATTCGATCAAGACCA 1589
 1661 TCAAAAGTGAAGTGGCTCTTAA 1683
 1590 TGCACAGGAGGCTGGCCTCTGA 1612

RESULT 11
 AL390197
 LOCUS
 DEFINITION
 Human DNA sequence from clone RP11-190F19 on chromosome 10,
 complete sequence.
 AL390197
 ACCESSION
 VERSION
 AL390197.19 GI:15387800
 KEYWORDS
 HTG.
 SOURCE
 Homo sapiens (human)
 ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 Tracey, A.
 REFERENCE
 1 (bases 1 to 188349)
 AUTHORS
 TITLE
 JOURNAL
 COMMENT
 Direct Submission
 Submitted (29-AUG-2001) Sanger Centre, Hinxton, Cambridgeshire,
 CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
 requests: clonerequest@sanger.ac.uk
 On Aug 31, 2001 this sequence version replaced gi:14596344.
 During sequence assembly data is compared from overlapping clones.
 Where differences are found these are annotated as variations
 together with a note of the overlapping clone name. Note that the
 variation annotation may not be found in the sequence submission
 corresponding to the overlapping clone, as we submit sequences with
 only a small overlap as described above.
 This sequence was finished as follows unless otherwise noted: all
 regions were either double-stranded or sequenced with an alternate
 chemistry or covered by high quality data (i.e., phred quality >=
 30); an attempt was made to resolve all sequencing problems, such
 as compressions and repeats; all regions were covered by at least
 one plasmid subclone or more than one M13 subclone; and the
 assembly was confirmed by restriction digest. The following
 abbreviations are used to associate primary accession numbers given
 in the feature table with their source databases: Em., EMBL; Sw.,
 SWISSPROT; Tr., TrEMBL; Wp., WORMPEP; Information on the WORMPEP
 database can be found at
 http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
 was generated from part of bacterial clone contigs of human
 chromosome 10, constructed by the Sanger Centre Chromosome 10
 Mapping Group. Further information can be found at
 http://www.sanger.ac.uk/HGP/Chr10
 RP11-190F19 is from the library RPCI-11.1 constructed by the group
 of Pieter de Jong. For further details see
 http://www.chori.org/bacpac/home.htm
 VECTOR: pBACE3.6
 This sequence is the entire insert of clone RP11-190F19 The true
 right end of clone RP11-481H12 is at 111825 in this sequence.

FEATURES
 Location/Qualifiers
 1..188349
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="10"
 /clone="RP11-190F19"
 /clone_lib="RPCI-11.1"
 22142..22187
 /notes="Sequence from overlapping clone RP11-481H12
 (AL133482). Assembly confirmed by restriction digest."

misc_feature
 ORIGIN

461 GCAGGCCAAACCCCTGCCAGATGGGCTACCTCTCCCGGCATTAAGCGGAGATCCAAAGT 520
 390 GCAGGCCAAACCCCTGCCAGATGGGCTACCTCTCCCGGCATTAAGCGGAGATCCAAAGT 449
 521 TCACCTGTGCTGCTCCCGACAGTTCAGGGAATTTCTGTGAATATAGTTCTGTGACT 580
 450 TTACCTGTGCTGCTCCCGACAGTTCAGGGAATTTCTGTGAATATAGTTCTGTGACT 509
 581 GCTATGTTGGGATGGCTTACTCTTACCAGAGGAAATGAATAGGACAGTCAACAGCATG 640
 510 GTTATGTCGGTGTGCTTACTCTTACCAGAGGAAATGAATAGGACAGTCAACAGCATG 569
 641 CGTGTCTTACTGGAATCCCACTCTCTTTCGAGGAGAAATTAACAATGTTTATGAGG 700
 570 CATGCTTTTACTGGAATCCCACTCTCTTTCGAGGAGAAATTAACAATGTTTATGAGG 629
 701 ATGCTGAACCCATCGGATTCGGGAACACAATTTCTGCAGAAACCCAGATCGGACCAAA 760
 630 ATGCAGAGACCCAGGATTCGAGGACACACTTCTGCAGAAACCCAGATCGGACCAAA 689
 761 AGCCCTGTGCTTTTAAAGTTACCAATGAAGTGAATGGAATGGAATGATGTCT 820
 690 AACCTGTGTTTCTCAAGGTGAACAGTGAAGGTGAATGGAATGGAATGATGTCT 749
 821 CAGCTGTGCTCAGCCAGGAGTGTCTTACCAGAGGAAAGCCCACTGAGCCATCAACCA 880
 750 CAGTCTGTCTCAGTGTGACACCCCTAACCCAGTGAAGGCTTCTGTGAGCTGTGATGG 809
 881 AGCTTCCGGGTTTGAATCTCTGTGAAAGACTGATAGCAGAGGAAATCAAGAGAA 940
 810 AGCTGCCAGGGTTTCAGTCTCTGCGGAAGACGAGGTAGTGAACACGACGATCAAGGTA 869
 941 TCTATGAGGCTTTAAGACAGCGCGGCGAGACCAATGACGAGGAAATCAAGAGAA 1000
 870 TCTATGAGGCTTTAAGACAGCGCGGCGAGACCAATGACGAGGAAATCAAGAGAA 929
 1001 CGCTGCTCTGACCTCTCTATCCCGGAGGCACTTCTGTGTGGGCGCTGATCCACC 1060
 930 CACTGCGGTGACCACTCTCTATCCCGGAGGCACTTCTGTGTGGGCGCTGATCCACC 989
 1061 CTTGCTGGTGTCTACTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1120
 990 CTTGCTGGTGTCTACTGCGAGCCCACTGTACCGACATAAACACCAAGCATTAAGATTG 1049
 1121 TGCTAGGAGCAGGACCTTGAAGAAAGAAATTTTCATGAGCAGAGCTTTAGGCTGCAGA 1180
 1050 TACTAGGAGATCAGGACCTTGAAGAAAGAAATTTTCATGAGCAGAGCTTTAGGCTGCAGA 1109
 1181 AGATATTCAGTACAGCCACTCAATGAAGAGATGAGATCCCAATGATATGAT 1240
 1110 AAATCTGAAGTACAGTCAAGTATAATGAAGAGATGAGATCCCAATGATATGAT 1169
 1241 TGCTCAAGTAAAGCAGGATGGTCACTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1300
 1170 TGCTCAAGTAAAGCAGGATGGTCACTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1229
 1301 TGCTGTGCTGATGGGCTCTTCCCTCTGCGAGTGAAGTGCACATCTCTGCTGGGGTG 1360
 1230 TATGTTTGGCCGAGCACCCTTCCCTCTGGAAGTGAAGTGCACATCTCTGCTGGGGTG 1289
 1361 TTACAGAAACAGAAAGGTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1420
 1290 TTACAGAAACAGAAAGGTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1349
 1421 ACATTTGTGCAACTCCCGCAACTCTATGACCAATGATGATGATGATGATGATGATGATG 1480
 1350 ACCCTTTGTGCAACTCCCGCAACTCTATGACCAATGATGATGATGATGATGATGATG 1409
 1481 CAGAAATCTTCAAGAACTGGGCAAGACCTGCGAGGTCAGTCTGAGAGCCCTCTGA 1540
 1410 CGGGGAACCTTCAAGAACTGGGCAAGACCTGCGAGGTCAGTCTGAGAGCCCTCTTAA 1469

Query Match 15.5%; Score 261.2; DB 9; Length 188349;
Best Local Similarity 93.8%; Pred. No. 1.6e-66;
Matches 272; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

QY 805 GAATAGTGTGATCTCTAGCGCTCTCAGCGCCAGAGCGTTGCTTACCCAGAGAAAGCCCC 864
DB 96119 GAATAGCACAAATTTATCTTTCTGTCTCCACAGACGTTGCTTACCCAGAGAAAGCCCC 96178

QY 865 ACTGAGCCATCAACCAAGCTTCCGGGGTTTGACTCTGTGGAAAGACTGAGATAGCAGAG 924
DB 96179 ACTGAGCCATCAACCAAGCTTCCGGGGTTTGACTCTGTGGAAAGACTGAGATAGCAGAG 96238

QY 925 AGGAAGATCAAGAGAAATCTATGAGGCTTTTAAGAGACACGCGGGGCAAGCACCCATGGCAG 984
DB 96239 AGGAAGATCAAGAGAAATCTATGAGGCTTTTAAGAGACACGCGGGGCAAGCACCCATGGCAG 96298

QY 985 GCGTCCCTCAGTCCTCGCTCTGACATCTCCATGCCCGAGGCCACTTCTGTGTGT 1044
DB 96299 GCGTCCCTCAGTCCTCGCTCTGACATCTCCATGCCCGAGGCCACTTCTGTGTGT 96358

QY 1045 GGGCGCTGATCCACCCCTGCTGGTGCTCACTGCTGCCACTGCACCGA 1094
DB 96359 GGGCGCTGATCCACCCCTGCTGGTGCTCACTGCTGCCACTGCACCGA 96408

RESULT 12
AC006097/c
LOCUS AC006097 157750 bp DNA linear HTG 30-AUG-2001
DEFINITION Homo sapiens chromosome 10 clone CIT987SK-1031G15 map 10q25.
ACCESSION AC006097
VERSION AC006097.1 GI:3962489
KEYWORDS HTG; HTGS PHASE2; HTGS_CANCELLED.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Smith,D.R.
Sequencing of Human Chromosome 10
Unpublished
2 (bases 1 to 157750)
Smith,D.R.
Direct Submission
Submitted (04-DEC-1998) Genome Therapeutics Corporation, 100 Beaver
Street, Waltham, MA 02154, USA
Note: Clone was sequenced in 1996 using Multiplex DNA Sequencing
Technology. Data may contain low quality sequence and BAC/Cosmid
vector sequences.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 157750: contig of 157750 bp in length.
FEATURES
source
1..157750
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="10"
/map="10q25"
/clone="CIT987SK-1031G15"

ORIGIN
Query Match 15.3%; Score 258; DB 2; Length 157750;
Best Local Similarity 93.1%; Pred. No. 1.5e-65;
Matches 270; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

QY 805 GAATAGTGTGATCTCTAGCGCTCTCAGCGCCAGAGCGTTGCTTACCCAGAGAAAGCCCC 864

DB 57685 GAATAGCACAAATTTATCTTTCTGTGTCTCCACAGACGTTGCTTACCCAGAGAAAGCCCC 57626
QY 865 ACTGAGCCATCAACCAAGCTTCCGGGGTTTGACTCTGTGGAAAGACTGAGATAGCAGAG 924
DB 57625 ACTGAGCCATCAACCAAGCTTCCGGGGTTTGACTCTGTGGAAAGACTGAGATAGCAGAG 57566

QY 925 AGGAAGATCAAGAGAAATCTATGAGGCTTTTAAGAGACACGCGGGGCAAGCACCCATGGCAG 984
DB 57565 AGGAAGATCAAGAGAAATCTATGAGGCTTTTAAGAGACACGCGGGGCAAGCACCCATGGCAG 57506

QY 985 GCGTCCCTCAGTCCTCGCTCTGACATCTCCATGCCCGAGGCCACTTCTGTGTGT 1044
DB 57505 GCGTCCCTCAGTCCTCGCTCTGACATCTCCATGCCCGAGGCCACTTCTGTGTGT 57446

QY 1045 GGGCGCTGATCCACCCCTGCTGGTGCTCACTGCTGCCACTGCACCGA 1094
DB 57445 GGGCGCTGATCCACCCCTGCTGGTGCTCACTGCTGCCACTGCACCGA 57396

RESULT 13
AX305372
LOCUS AX305372 397 bp DNA linear PAT 11-DEC-2001
DEFINITION Sequence 123 from Patent WO0188188.
ACCESSION AX305372
VERSION AX305372.1 GI:17644923
KEYWORDS Mus musculus (house mouse)
SOURCE Mus musculus
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1
REFERENCE Ishikawa,K., Asai,S., Takahashi,Y., Nagata,T. and Ishii,Y.
AUTHORS Method for examining ischemic conditions
TITLE Patent: WO 0188188-A 123 22-NOV-2001;
JOURNAL School Juridical Person Nihon University (JP)
FEATURES
source
1..397
/organism="Mus musculus"
/mol_type="unassigned DNA"
/db_xref="taxon:10090"

ORIGIN
Query Match 14.3%; Score 241; DB 6; Length 397;
Best Local Similarity 84.4%; Pred. No. 8.7e-61;
Matches 271; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 1363 ACAGAAACAGAGAAAGGGTCCCGCCAGCTCCTGGATGCCAAAGTCAAGCTGATTGCCAAC 1422
DB 13 ACAGAAACAGAGAAAGGGTCCCGCCAGCTCCTGGATGCCAAAGTCAAGCTGATTGCCAAC 72

QY 1423 ACTTGTGCACTCCCGCAACTCTATGACACATGATTGATGACAGTATGATCTGTGCA 1482
DB 73 CTTTGTGCACTCCCGCAACTCTATGACACATGATTGATGATGATGATGATGATG 132

QY 1483 GGAAATCTTCAGAAACCTGGGCAAGACACCTGCCAGGGTGACTCTGGAGGGCCCTGACC 1542
DB 133 GGAACTTCAGAAACCTGGGCAAGACACCTGCCAGGGTGACTCTGGAGGGCCCTGACC 192

QY 1543 TGTGAGAGGAGCGCACTTACTTACTTATGGGATAGTAGCTGGGGCTGGAGTGTGAG 1602
DB 193 TGTGAGAGGAGTGAAGTCTTACTTACTTACGGGATTTGAAGCTGGGGCCAGGAATGTGGG 252

QY 1603 AAGAGCCAGGGGTCTACACCCCAAGTTACCAATTTCTGAATTTGGATCAAGCCACCATC 1662
DB 253 AAGAGCCAGGAGTCTACACTCAAGTCCCAAGTTCTGATTTGATTAAGACCAACCATG 312

QY 1663 AAAAGTGAAGTGGCTTTCTAA 1683
DB 313 CACAGGGAGGCTGGGCTCTGA 333

RESULT 14

AC115771

LOCUS
DEFINITION
AC115771
VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS

AC115771 204214 bp DNA linear HTG 18-JUN-2003
Mus musculus clone RP23-116D4, WORKING DRAFT SEQUENCE, 11 ordered
pieces.
AC115771
HTG: HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 204214)
Birken, B., Nusbaum, C. and Lander, E.
Mus musculus, clone RP23-116D4
Unpublished
2 (bases 1 to 204214)

Birken, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,
Anderson, S., Barina, N., Bastien, V., Bloom, T., Boguski, M.,
Bouckgeert, B., Brown, A., Camarata, J., Campopiano, A., Chang, J.,
Chazaro, B., Choepel, Y., Collings, M., Collins, S., Collymore, A.,
Cook, A., Cooke, P., DeArelano, K., Dewar, K., Diaz, J.S., Dodge, S.,
Faro, S., Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S.,
Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
Kamat, A., Karatas, A., Kelle, C., Lacroque, K., Lamazares, R.,
Lander, E., Lehoucq, J., Levine, R., Lindblad-Toh, K., Liu, G.,
Maclean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C.,
McCarthy, M., McEwan, P., McKernan, K., Meldrum, J., Meneus, L.,
Mihova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R.,
Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D.,
Olivier, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,
Raymond, C., Retta, R., Riebeck, M., Riley, R., Rise, C., Rogov, P.,
Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupack, R.,
Seaman, S., Severy, P., Spencer, B., Stange-Thoman, N., Stojanovic, N.,
Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
Topham, K., Travers, M., Travis, N., Trigglio, J., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G.,
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE
JOURNAL
REFERENCE
AUTHORS

Direct Submission
Submitted (22-MAR-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 204214)
Birken, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N.,
Anderson, M., Arachchi, H.M., Barina, N., Bastien, V., Bloom, T.,
Boguski, M., Bouckgeert, B., Camarata, J., Chang, J., Choepel, Y.,
Collins, M., Cooke, P., Corra, B., Dorris, L., Erickson, J., Faro, S.,
Ferreira, P., FitzGerald, M., Gage, D., Galagan, J., Gardyna, S.,
Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagos, B.,
Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
Kamat, A., Karatas, A., Kelle, C., Lander, E., Levine, R.,
Lindblad-Toh, K., Liu, X., Lui, A., Mabbitt, R., Maclean, C.,
Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M.,
Meldrum, J., Meneus, L., Mihova, T., Mienga, V., Murphy, T., Naylor, J.,
Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P.,
O'Neill, D., Olivier, J., Peterson, K., Phunkhang, P., Pierre, N.,
Rachupka, A., Ramasamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P.,
Roman, J., Schauer, S., Schupack, R., Seaman, S., Severy, P., Smith, C.,
Spencer, B., Stange-Thoman, N., Stojanovic, N., Stubbs, M.,
Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M.,
Vassiliev, H., Venkataraman, V.S., Viel, R., Vo, A., Wilson, B., Wu, X.,
Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
Direct Submission
Submitted (18-JUN-2003) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jun 18, 2003 this sequence version replaced gi:21532088.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RN/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIRB
Web site: http://www-seq.wi.mit.edu

Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L23373
Center clone name: 116.D.4
----- Summary Statistics
Sequencing vector: Plasmid, n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 201691 bases at least Q40
Consensus quality: 202492 bases at least Q30
Consensus quality: 202921 bases at least Q20
Insert size: 210000; agarose- ϕ
Insert size: 203214; sum-of-contigs
Quality coverage: 12.4 in Q20 bases; agarose- ϕ
Quality coverage: 12.9 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 11 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

* 70662: contig of 70662 bp in length
* 70663: gap of 100 bp
* 70664: contig of 671 bp in length
* 71433: gap of 100 bp
* 71434: contig of 838 bp in length
* 72371: gap of 100 bp
* 72372: contig of 1449 bp in length
* 72472: gap of 100 bp
* 73221: contig of 2445 bp in length
* 74021: gap of 100 bp
* 76456: contig of 10038 bp in length
* 76566: gap of 100 bp
* 86604: contig of 16553 bp in length
* 86704: gap of 100 bp
* 103257: gap of 100 bp
* 103257: gap of 100 bp
* 103257: gap of 100 bp
* 117661: contig of 14305 bp in length
* 117662: gap of 100 bp
* 135962: contig of 18201 bp in length
* 135963: gap of 100 bp
* 135963: contig of 45329 bp in length
* 181392: gap of 100 bp
* 181392: contig of 22723 bp in length.
* 181492: contig of 22723 bp in length.

FEATURES
source
1. 204214
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/clone="RP23-116D4"
/clone_lib="RPCI-23 Female Mouse BAC"
1. 70662
/note="assembly_fragment"
clone_end:SP6
vector_side:left
70663. 71433
/note="assembly_fragment"
71534. 72371
/note="assembly_fragment"
72472. 73221
/note="assembly_fragment"
74021. 76456
/note="assembly_fragment"
76566. 86603
/note="assembly_fragment"
86704. 103256
/note="assembly_fragment"
103357. 117661
/note="assembly_fragment"
117762. 135962
/note="assembly_fragment"

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature


```

misc_feature 136063..181391
             /note="assembly_fragment"
misc_feature 181492..204214
             /note="assembly_fragment"
             clone_end=T7
             vector_side:right"

ORIGIN
Query Match      9.9%; Score 167.4; DB 2; Length 204214;
Best Local Similarity 78.2%; Pred. No. 3.2e-38;
Matches 201; Conservative 0; Mismatches 56; Indels 0; Gaps 0;

QY 838 GAGCTTGCTTACCCAGAGAAAGCCCACTGAGCCATCAACCAAGCTTCGGGGGTTTGAC 897
Db 142187 GACACCCCTAACCCAGTGGAAAGCCCTCTCGAGCGCTGATGAGAGCTGCCAGGGTTGAG 142246

QY 898 TCTGTGGAAAGACTAGATAGCAGAGAGAGATCAAGAGATCTATGAGAGCTTTAAG 957
Db 142247 TCTGTGGGGAAGACGGAGTAGTGTGAACACGACGATCAAGCGTATCTACGGGGGCTTTAAG 142306

QY 958 AGCAGCGGGGCGAAGACACCCATCGCAGGGGTCCTCCAGTCTCGTGCCTCTGACCATC 1017
Db 142307 AGCAGCAGGCGAAGACCCGTCGAGGTGTCTCTCGACACCTCACTGCCGTTGACCAC 142366

QY 1018 TCATGCCCCAGGGCCACTTCTGTGTGGGGGCTGATCCACCCCTGCTGGTGCTCACT 1077
Db 142367 TCCATGCCCCAGGCGACCTTCTGTGTGGGGGCGCCCTGATCCACCCCTGCTGGTGCTCACT 142426

QY 1078 GCTGCCCACTGCACCGA 1094
Db 142427 GCAGCCACTGTACCGA 142443

```

```

RESULT 15
AC106236/c
LOCUS      Rattus norvegicus clone CH230-131J10, WORKING DRAFT SEQUENCE, 2
unordered pieces.
ACCESSION  AC106236
VERSION     AC106236.5 GI:30579797
KEYWORDS   HTG; HTGS_PRAISE1; HTGS DRAFT; HTGS_FULLTOP.
SOURCE     Rattus norvegicus
ORGANISM   Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE  1 (bases 1 to 224208)
            Muzny,D.,Marie., Metzker,M.,Lee., Abramson,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denison,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gregorogis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamill,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,E., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpachy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,D., Li,Z., Liu,J.,
            Liu,J., Liu,W., London,P., Longacre,S., Lopez,J.,
            Lorenshewa,L., Loulseeq,H., Lozada,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mashiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Parks,K.,
            Paokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkuch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puzos,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Rose,R., Ruiz,S.J.,
            Rivers,C., Rodkey,I., Rojas,A., Rose,M., Richards,S., Riggs,F.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajls,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Taber,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villaseana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,D., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
            Direct Submission
            Unpublished
            2 (bases 1 to 224208)
            Worley,K.C.
            Direct Submission
            Submitted (12-JAN-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
            3 (bases 1 to 224208)
            Rat Genome Sequencing Consortium.
            Direct Submission
            Submitted (13-MAY-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
            On May 13, 2003 this sequence version replaced gi:24942489.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            ----- Genome Center of Medicine
            Center: Baylor College of Medicine
            Web site: http://www.hgsc.bcm.tmc.edu/
            Contact: hgsc-help@bcm.tmc.edu
            ----- Project Information
            Center project name: GKCC
            Center clone name: CH230-131J10
            ----- Summary Statistics
            Assembly program: Atlas 3.0;
            Consensus quality: 213390 bases at least Q40
            Consensus quality: 215800 bases at least Q30
            Consensus quality: 216989 bases at least Q20
            Estimated insert size: 222031; sum-of-contigs estimation
            Quality coverage: 8x in Q20 bases; sum-of-contigs estimation
            -----
            * NOTE: Estimated insert size may differ from sequence length
            (see http://www.hgsc.bcm.tmc.edu/docs/Genbank draft data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            consists of 2 contigs. The true order of the pieces
            is not known and their order in this sequence record is
            arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.

```


* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 196130: contig of 196130 bp in length
* 196131 196230: gap of unknown length
* 196231 224208: contig of 27978 bp in length.
Location/Qualifiers
1..224208
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-131J10"
2668..3129
/note="clone boundary"
clone_end:Sp6
site:EcoRI
end_sequence:BH346758"
200382..201926
/note="wgs_contig"
ORIGIN
Query Match 9.6%; Score 161; DB 2; Length 224208;
Best Local Similarity 76.7%; Pred No. 2.8e-36;
Matches 197; Conservative 0; Mismatches 60; Indels 0; Gaps 0;
QY 838 GACGTGCTACCCAGAGAAAGCCCACTGAGCCATCAACCAAGCTTCGGGGTTTCAC 897
Db 71048 GATGCTGTACCCATTGGCGACCTTCAGGAACCTGTGATGGAGTCCAGGATTCGAC 70989
QY 898 TCCTGTGGAAGACTGAGATGCAGAGAGAGATCAAGAGATCTATGGAGGCTTTAAG 957
Db 70989 TCCTGCGGGAAGACAGAGATGACTGAACACGCGGTCAAGCGCATCTACGGGGCTTTAAG 70929
QY 958 AGCAGCGCGGCAAGCACCATCGGAGCGTCCTCCAGTCCCTGCTGCTGTGACCATC 1017
Db 70928 AGCAGAGCGGCAAGCACCCTGGCAGGTGTCCTGCGACCTCATTGCGCTGACCACC 70869
QY 1018 TCCATGCCCCAGGGCCACTTCTGTGTGGGGGCTGATCCACCCCTGCTGGTGTCTACT 1077
Db 70868 TCCATGCCCCAGGGCCACTTCTGTGTGGGGGTTCCCTGATTCAACCCCTGCTGGTGTCTACT 70809
QY 1078 GCTGCCCACTGCACCGA 1094
Db 70808 GCAGCCCACTGTACCGA 70792

Search completed: May 25, 2004, 09:00:34
Job time : 4483.5 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 25, 2004, 04:19:43 ; Search time 471 Seconds
(without alignments)

15179.867 Million cell updates/sec

Title: US-09-912-559-2

Perfect score: 1683

Sequence: 1 atgtttccaggtgtctga.....aaagtgaagcttcttaa 1683

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_29Jan04:*

- 1: Geneseqn1980s:*
- 2: Geneseqn1990s:*
- 3: Geneseqn2000s:*
- 4: Geneseqn2001as:*
- 5: Geneseqn2001bs:*
- 6: Geneseqn2002s:*
- 7: Geneseqn2003as:*
- 8: Geneseqn2003bs:*
- 9: Geneseqn2003cs:*
- 10: Geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1683	100.0	1683	6	AAL45697 Human blo
2	1679.8	99.8	1683	6	AAL45696 Human blo
3	1679.8	99.8	3008	6	ABN95170 Gene #166
4	1679.8	99.8	3008	7	ACC51204 Human Plk
5	1675	99.5	3623	3	AAC76693 Human ORF
6	526.4	31.3	617	3	AAA44763 Human sec
7	368.4	21.9	428	8	ACH21030 Human adu
8	241	14.3	397	6	AB192821 Mouse isc
C 9	120.2	7.1	451	4	ABA58323 Human fce
C 10	120.2	7.1	451	4	AA137947 Probe #66
C 11	120.2	7.1	451	4	AAK32096 Human bon
C 12	120.2	7.1	451	4	AAK06420 Human bra
C 13	120.2	7.1	451	4	ABS31791 Human liv
C 14	120.2	7.1	451	6	ABS06863 Human gen
C 15	117	7.0	117	4	ABA70919 Human fce
C 16	117	7.0	117	4	AA151106 Probe #19
C 17	117	7.0	117	4	AAK45153 Human bon
C 18	117	7.0	117	4	AAK19188 Human liv
C 19	117	7.0	117	4	ABS44824 Human liv
C 20	117	7.0	117	6	ABS19403 Human gen
C 21	114.4	6.8	2033	2	AAQ63951 Hepatocyt
C 22	114.4	6.8	2036	6	AB165242 Lung canc
C 23	114.4	6.8	2036	6	ABN95114 Gene #161

ALIGNMENTS

RESULT 1

AAL45697
ID AAL45697 standard; DNA; 1683 BP.

XX AAL45697;

XX AAL45697;

XX AAL45697;

DT 13-JUN-2002 (first entry)

XX Human blood coagulation factor VII activating protease mutant DNA.

XX Human; blood coagulation factor VII activating protease; FSAP;

KW single-chain plasminogen activator; bleeding disorder; haematological;

KW haemostatic; mutant; gene; ds.

XX Homo sapiens.

OS Synthetic.

XX Synthetic.

XX Synthetic.

XX Synthetic.

XX Synthetic.

XX Synthetic.

XX Synthetic.

XX Synthetic.

XX Synthetic.

XX Synthetic.

XX Synthetic.

XX Synthetic.

XX Synthetic.

XX Synthetic.

XX Synthetic.

XX Synthetic.

XX Synthetic.

XX Synthetic.

XX Synthetic.

XX Synthetic.

XX Synthetic.

XX Synthetic.

XX Synthetic.

XX Synthetic.

XX Synthetic.

XX Synthetic.

XX Synthetic.

XX Synthetic.

XX Synthetic.

XX Synthetic.

XX Synthetic.

XX Synthetic.

XX Synthetic.

XX Synthetic.

XX Synthetic.

XX Synthetic.

XX Synthetic.

XX Synthetic.

XX Synthetic.

XX Synthetic.

QY 1441 CAACCTATGACACATGATTGATGACAGATATGATCTGTGACGAAATCTTCAGAAACCT 1500
 Db 1441 CNACTCTATGACACATGATTGATGACAGATATGATCTGTGACGAAATCTTCAGAAACCT 1500
 QY 1501 GGGCAAGACACCTGCCAGGTGACTCTGAGAGGCCCTGACCTGTGAGAGGACGGCACC 1560
 Db 1501 GGGCAAGACACCTGCCAGGTGACTCTGAGAGGCCCTGACCTGTGAGAGGACGGCACC 1560
 QY 1561 TACTACCTCTATGGGATAGTGAAGCTGGGGCTGGAGTGTGAGAGAGGCCAGGGGCTTAC 1620
 Db 1561 TACTACCTCTATGGGATAGTGAAGCTGGGGCTGGAGTGTGAGAGAGGCCAGGGGCTTAC 1620
 QY 1621 ACCCAAGTTACCAATTCCTGAATTTGGATCAAAAGCCACCAATCAAAAGTGAAGTGGCTTC 1680
 Db 1621 ACCCAAGTTACCAATTCCTGAATTTGGATCAAAAGCCACCAATCAAAAGTGAAGTGGCTTC 1680
 QY 1681 TAA 1683
 Db 1681 TAA 1683

RESULT 3

ABN95170
 ID ABN95170 standard; DNA; 3008 BP.

XX AC ABN95170;

XX 13-AUG-2002 (first entry)

XX Gene #1668 used to diagnose liver cancer.

XX Gene; liver cancer; ds; hepatocellular carcinoma; hepatotrophic;
 XX metastatic liver tumor; cytostatic; expression profiler; disease state;
 XX disease progression; drug toxicity; drug efficacy; drug metabolism.

XX Homo sapiens.

XX W0200229103-A2.

XX 11-APR-2002.

XX 02-OCT-2001; 2001WO-US030589.

XX 02-OCT-2000; 2000US-0237054P.

XX (GENE-) GENE LOGIC INC.

XX Horne D, Alvares C, Peres-Da-Silva S, Vockley JG;

XX WPI; 2002-426119/45.

XX Diagnosing and detecting the progression of liver cancer, hepatocellular
 XX carcinoma or metastatic liver tumor in a patient, involves detecting the
 XX level of expression of two or more genes in a liver tissue sample.

XX Claim 1; SEQ ID NO 1668; 298pp; English.

XX The invention relates to a novel method for diagnosing and detecting the
 XX progression of liver cancer, hepatocellular carcinoma or metastatic liver
 XX tumor in a patient, and differentiating metastatic liver cancer from
 XX hepatocellular carcinoma in a patient, involving detecting the level of
 XX expression of two or more genes represented in ABN93503-ABN97455 in a
 XX tissue sample. The method of the invention has hepatotrophic, and
 XX cytostatic activity. The method is useful for diagnosing and detecting
 XX the progression of liver cancer, hepatocellular carcinoma and metastatic
 XX liver carcinoma in a patient. The method is useful for identifying
 XX expression profiles which serve as useful diagnostic markers as well as
 XX markers that can be used to monitor disease states, disease progression,
 XX drug toxicity, drug efficacy and drug metabolism. Note: The sequence data
 XX for this patent did not form part of the printed specification, but was
 XX obtained in electronic format directly from WIPO at
 XX ftp.wipo.int/pub/published_pct_sequences

SQL Sequence 3008 BP; 791 A; 837 C; 703 G; 677 T; 0 U; 0 Other;
 Query Match 99.8%; Score 1679.8; DB 6; Length 3008;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 1681; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 ATGTTTGCAGAGTGTCTGATCTCCATGTTCTGTGTTAATGCTCTGGTGGGAAAGACA 60
 Db 97 ATGTTTGCAGAGTGTCTGATCTCCATGTTCTGTGTTAATGCTCTGGTGGGAAAGACA 156
 QY 61 GCCTGTGGTTCCTCCCTGATGTCCTTTATTTGGAAGCCTGGACCCAGATGGACCCCTGAC 120
 Db 157 GCCTGTGGTTCCTCCCTGATGTCCTTTATTTGGAAGCCTGGACCCAGATGGACCCCTGAC 216
 QY 121 CAGTATGATTACAGCTACGAGGATTAATAATCAGAGAGAGAACACACAGTAGCACACTTACC 180
 Db 217 CAGTATGATTACAGCTACGAGGATTAATAATCAGAGAGAGAACACACAGTAGCACACTTACC 276
 QY 181 CATGCTGAGAATCTGACTGTGATCTACACTGAGGACCAAGCTGATCTCATGCGCCCAAC 240
 Db 277 CATGCTGAGAATCTGACTGTGATCTACACTGAGGACCAAGCTGATCTCATGCGCCCAAC 336
 QY 241 CCCTGTGAACACGGTGGGACTGCTCGTCCATGGGAGCACCTTTCACATGAGCTGGCTG 300
 Db 337 CCCTGTGAACACGGTGGGACTGCTCGTCCATGGGAGCACCTTTCACATGAGCTGGCTG 396
 QY 301 GCTCCTTTCTCTGGGATTAAGTGTGAGAAAGTGAATAATGCTGCAAGGACCAACCCATGT 360
 Db 397 GCTCCTTTCTCTGGGATTAAGTGTGAGAAAGTGAATAATGCTGCAAGGACCAACCCATGT 456
 QY 361 GGCGGGGCCAATGTCTCATTACCACAGAGTCTCCCTACTACCGCTGTGTGTAACAC 420
 Db 457 GGCGGGGCCAATGTCTCATTACCACAGAGTCTCCCTACTACCGCTGTGTGTAACAC 516
 QY 421 CTTTACACAGTCCCGAGTCTCCCAAGTGTCTCTGTATGACGAGCCAAACCCCTGGCAG 480
 Db 517 CTTTACACAGTCCCGAGTCTCCCAAGTGTCTCTGTATGACGAGCCAAACCCCTGGCAG 576
 QY 481 AATGGGCTTACTGCTCCCGGCATAAGCGGAGATCCCAAGTTCACTGCTGCTGCTCCGAC 540
 Db 577 AATGGGCTTACTGCTCCCGGCATAAGCGGAGATCCCAAGTTCACTGCTGCTGCTCCGAC 636
 QY 541 CAGTTCAAGGGGAAATCTGTGAAATAGTTTCTGATGACTGCTGATGTTGGGATGGCTAC 600
 Db 637 CAGTTCAAGGGGAAATCTGTGAAATAGTTTCTGATGACTGCTGATGTTGGGATGGCTAC 696
 QY 601 TCTTACCGAGGAAATGATAGGACAGTCAACAGCATGCTGCTCCCTTTACTGGAATCC 660
 Db 697 TCTTACCGAGGAAATGATAGGACAGTCAACAGCATGCTGCTCCCTTTACTGGAATCC 756
 QY 661 CACTCTCTTTGCGAGGAAATTAACAATGTTTATGAGGATGCTGAAACCCATGGGATT 720
 Db 757 CACTCTCTTTGCGAGGAAATTAACAATGTTTATGAGGATGCTGAAACCCATGGGATT 816
 QY 721 GGGGAACAATTTCTGCAGAAACCCAGATGCGGACGAAAGCCCTGGTGTCTTTATAA 780
 Db 817 GGGGAACAATTTCTGCAGAAACCCAGATGCGGACGAAAGCCCTGGTGTCTTTATAA 876
 QY 781 GTTACCAATGACAGGTGAAATGGAATACTGTGATGCTCAGCTGCTCAGCCAGGAC 840
 Db 877 GTTACCAATGACAGGTGAAATGGAATACTGTGATGCTCAGCTGCTCAGCCAGGAC 936
 QY 841 GTTGTCTACCGAGGAAAGCCCACTGAGCCATCAACAGCTTCCGGGTTTGAATCC 900
 Db 937 GTTGTCTACCGAGGAAAGCCCACTGAGCCATCAACAGCTTCCGGGTTTGAATCC 996
 QY 901 TGTGAAAGACTGAGATAGCAGAGGAGGAAATCAAGAGAAATCTATGGAGGCTTTAAGAGC 960
 Db 997 TGTGAAAGACTGAGATAGCAGAGGAGGAAATCAAGAGAAATCTATGGAGGCTTTAAGAGC 1056
 QY 961 ACGGGGCGGACAGCCCATGACAGGCTCCCTCGCTGCTGCTGCTGCTGCTGCTGCTGCT 1020
 Db 1057 ACGGGGCGGACAGCCCATGACAGGCTCCCTCGCTGCTGCTGCTGCTGCTGCTGCTGCT 1116

QY 1021 ATGCCCGAGGCGCACTTCTGTGGTGGGCGCTGATCCACCCCTGCTGGGTGCTCACTGCT 1080
Db ATGCCCGAGGCGCACTTCTGTGGTGGGCGCTGATCCACCCCTGCTGGGTGCTCACTGCT 1176
QY 1081 GCCCACTGCACCGCATATAAAACAGACATCTAAAGTGTGTCTAGGGACACAGGACCTG 1140
Db GCCCACTGCACCGCATATAAAACAGACATCTAAAGTGTGTCTAGGGACACAGGACCTG 1236
QY 1141 AAGAAAGAAATTTCTATGACAGAGCTTTAGGCTGAGAGATATTTCAAGTACAGCCAC 1200
Db AAGAAAGAAATTTCTATGACAGAGCTTTAGGCTGAGAGATATTTCAAGTACAGCCAC 1296
QY 1201 TACAATGAAGAGATGAGATTTCCCAACAATGATATTCATTTGCTCAAGTTAAAGCCAGTG 1260
Db TACAATGAAGAGATGAGATTTCCCAACAATGATATTCATTTGCTCAAGTTAAAGCCAGTG 1356
QY 1261 GATGGTCACTGTCTGTAGAAATCAAAATACGTGAAGACTGTGTGCTTGGCTGTGAGGTGC 1320
Db GATGGTCACTGTCTGTAGAAATCAAAATACGTGAAGACTGTGTGCTTGGCTGTGAGGTGC 1416
QY 1321 TTTCCCTCTGGAGTGAGTGCCACATCTCTGGCTGGGTGTTACAGAAACAGGAAAGGG 1380
Db TTTCCCTCTGGAGTGAGTGCCACATCTCTGGCTGGGTGTTACAGAAACAGGAAAGGG 1476
QY 1381 TCCGCGCAGCTCTGGATGCCAAAGTCAAGCTGATTTGCCAACACTTTGTGCAACTCCCGC 1440
Db TCCGCGCAGCTCTGGATGCCAAAGTCAAGCTGATTTGCCAACACTTTGTGCAACTCCCGC 1536
QY 1441 CAATCTATGACCAATGATTTGATGACAGTATGATCTGTCAGGAATCTTCAGAACCT 1500
Db CAATCTATGACCAATGATTTGATGACAGTATGATCTGTCAGGAATCTTCAGAACCT 1596
QY 1501 GGGCAAGACACTGCCAGGCTGACTCTGGAGGCCCCCTGACCTGTGAGAGGACGGCACC 1560
Db GGGCAAGACACTGCCAGGCTGACTCTGGAGGCCCCCTGACCTGTGAGAGGACGGCACC 1656
QY 1561 TACTACGTCTATGGATAGTACGTGGGCTGAGTGTGAGAGACCCAGGGGTCTAC 1620
Db TACTACGTCTATGGATAGTACGTGGGCTGAGTGTGAGAGACCCAGGGGTCTAC 1716
QY 1621 ACCCAAGTTACCAATTCCTGAATTTGGATCAAGCCACCATCAAAAGTGAAGTGGCTTC 1680
Db ACCCAAGTTACCAATTCCTGAATTTGGATCAAGCCACCATCAAAAGTGAAGTGGCTTC 1776
QY 1681 TAA 1683
Db TAA 1779

RESULT 4
ACCS1204
ID ACCS1204 standard; cDNA; 3008 BP.
AC ACCS1204;
XX
XX
DE 16-JUN-2003 (first entry)
XX Human Plk-1 related cDNA sequence hmft-0306 SEQ ID NO:89.
XX Human; hepatoblastoma; cancer detection probe; cancer; detection;
KW hepatocellular carcinoma; hereditary non-polyposis colorectal cancer;
KW desmoid tumour; anaplastic thyroid carcinoma; Wilm's tumour; tumour;
XX Plk-1; polo-like kinase-1; gene; ss.
OS Homo sapiens.
XX
PN WO2003018807-A1.
XX
PD 06-MAR-2003.
XX
PF 26-AUG-2002; 2002WO-JP0008580.
XX

PR 24-AUG-2001; 2001JP-00255225.
XX (HISM) HISAMITSU PHARM CO LTD.
PA (CHIB-) CHIBA PREFECTURE.
XX
XX Nakagawara A;
PI
XX WPI; 2003-268424/26.
XX
XX Nucleic acid sequences differently expressed between hepatoblastoma and
PT normal liver tissue, are useful for cancer detection and diagnosis.
XX
XX Claim 4; Page 142-144; 180pp; Japanese.
XX
XX The present invention describes nucleic acid sequences (I) having a
CC different degree of expression in hepatoblastoma from their expression in
CC normal liver tissue. ACCS1116 to ACCS1219 represents specifically claimed
CC examples of (I). Also described: (1) nucleic acids stringently
CC hybridising to (I); (2) cancer detection probes containing one or more of
CC 104 listed sequences (II), see ACCS1116 to ACCS1219, including the 79 (I,
CC see ACCS1116 to ACCS1194), or partial sequences derived from them; (3)
CC PCR primers for cancer detection based on sequences (II); (4) marker
CC proteins for cancer detection, encoded by (II); (5) diagnostic reagents
CC for cancer diagnosis, containing (II) or their partial sequences. The
CC nucleic acid sequences are useful in the detection and diagnosis of
CC cancers including liver, colon, breast, kidney, bladder, ovary and
CC thyroid cancer, especially for hepatoblastoma, hepatocellular carcinoma,
CC hereditary non-polyposis colorectal cancer, desmoid tumour, anaplastic
CC thyroid carcinoma and Wilm's tumour. They are also used as markers for
CC predicting the prognosis of these tumours. ACCS1220 to ACCS1233 represent
CC PCR primers used in the exemplification of the present invention. The
CC nucleic acid sequences given in ACCS1116 to ACCS1219 are related to human
CC Plk-1 (polo-like kinase-1), which is located on chromosome 16p12
XX
XX Sequence 3008 BP; 791 A; 837 C; 703 G; 677 T; 0 U; 0 Other;
Query Match 99.8%; Score 1679.8; DB 7; Length 3008;
Best Local Similarity 99.9%; Pred No. 0;
Matches 1681; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 ATGTTTGGCAGATGTCGATCTCCATGTTCTGTGTTAAATGGCTCTGGTGGAAAGACA 60
Db 97 ATGTTTGGCAGATGTCGATCTCCATGTTCTGTGTTAAATGGCTCTGGTGGAAAGACA 156
QY 61 GCCTGTGGGTTCTCCCTGATGTTCTTATTGGAAGCCCTGGACCAGCTGGACCCCTGAC 120
Db 157 GCCTGTGGGTTCTCCCTGATGTTCTTATTGGAAGCCCTGGACCAGCTGGACCCCTGAC 216
QY 121 CAGTATGATTACAGCTACGAGGATTATATATCAGGAAGAGAACACCACTAGCACACTTACC 180
Db 217 CAGTATGATTACAGCTACGAGGATTATATATCAGGAAGAGAACACCACTAGCACACTTACC 276
QY 181 CATGCTGAGAACTCTGACTGTGTACTACACTGAGGACCAAGCTGATCCATGCCAGCCCAAC 240
Db 277 CATGCTGAGAACTCTGACTGTGTACTACACTGAGGACCAAGCTGATCCATGCCAGCCCAAC 336
QY 241 CCCTGTGAACACGGTGGGACTGCGCTCGCTCCATGGAGACACCTTCACATGACGTGCGCTG 300
Db 337 CCCTGTGAACACGGTGGGACTGCGCTCGCTCCATGGAGACACCTTCACATGACGTGCGCTG 396
QY 301 GTCCTTTCTCTGGGAAATAGTGTGAGAAAGTGCAGAAATACCTGCGAGGACCAACCATCT 360
Db 397 GTCCTTTCTCTGGGAAATAGTGTGAGAAAGTGCAGAAATACCTGCGAGGACCAACCATCT 456
QY 361 GGCCGGGGCCCAATGTCCTATTACCCAGAGTCTCCCTACTACCGTGTGTCTGTAAACAC 420
Db 457 GGCCGGGGCCCAATGTCCTATTACCCAGAGTCTCCCTACTACCGTGTGTCTGTAAACAC 516
QY 421 CCTTACACAGGTCCAGCTGCTCCCAAGTGGTTCTGTATGAGGCGCAACCCCTGCCAG 480
Db 517 CCTTACACAGGTCCAGCTGCTCCCAAGTGGTTCTGTATGAGGCGCAACCCCTGCCAG 576
QY 481 AATGGGGCTACCTGCTCCCGGATTAAGCGGAGATCCAAAGTTACCTGTGCTGTCCCGAC 540

577 AATGGGGTACCTGCTCCCGCATAAAGCGAGATCAAGTTCACCTGTGCTGCTCCGAC 636
 541 CAGTTCAAGCGGAAATCTGTGAATAGTCTGTGATGCTGTGATGCTGTGCGGATGGCTAC 600
 637 CAGTTCAAGCGGAAATCTGTGAATAGTCTGTGATGCTGTGATGCTGTGCGGATGGCTAC 696
 601 TCTTACCGAGGAAATGAATAGACAGTCAACAGCATGCGTGCTGCTTTACTGGAATCC 660
 697 TCTTACCGAGGAAATGAATAGACAGTCAACAGCATGCGTGCTGCTTTACTGGAATCC 756
 661 CAGCTCTCTTGTGAGAGAAATTAACAATGTTTATGGAGATGCTGAAACCCATGGGATT 720
 757 CAGCTCTCTTGTGAGAGAAATTAACAATGTTTATGGAGATGCTGAAACCCATGGGATT 816
 721 GGGGAACACAAATTTCTGACAGAAACCCAGATGCGGACGAAAGCCCTGCTTTTATAA 780
 817 GGGGAACACAAATTTCTGACAGAAACCCAGATGCGGACGAAAGCCCTGCTTTTATAA 876
 781 GTTACCAATGACAGGTAATGGGAATAGTGTCTCAGCTGCTCAGCCAGGAC 840
 877 GTTACCAATGACAGGTAATGGGAATAGTGTCTCAGCTGCTCAGCCAGGAC 936
 841 GTTGCCTTACCAGAGGAAACCCACTGAGCCATCAACCAAGCTTCCGGGGTTTGAATCC 900
 937 GTTGCCTTACCAGAGGAAACCCACTGAGCCATCAACCAAGCTTCCGGGGTTTGAATCC 996
 901 TGTGGAAGACTGATAGACAGAGAGAAATCAAGAGATCTATGAGGCTTTAAGAC 960
 997 TGTGGAAGACTGATAGACAGAGAGAAATCAAGAGATCTATGAGGCTTTAAGAC 1056
 961 ACGGCGGCAAGCAACCATGCGAGGCTGCTCCAGTCTGCTGCTGCTGCTGCAATCTCC 1020
 1057 ACGGCGGCAAGCAACCATGCGAGGCTGCTCCAGTCTGCTGCTGCTGCTGCAATCTCC 1116
 1021 ATGCCCCAGGCGCACTTCTGTGTGGGCGCTGATCCACCCCTGCTGGGTGCTCACTGCT 1080
 1117 ATGCCCCAGGCGCACTTCTGTGTGGGCGCTGATCCACCCCTGCTGGGTGCTCACTGCT 1176
 1081 GCCCACTGCAACGACATAAAACAGACATCTAAAGTGTGTGTAGGGACACAGGACCTG 1140
 1177 GCCCACTGCAACGACATAAAACAGACATCTAAAGTGTGTGTAGGGACACAGGACCTG 1236
 1141 AAGAAAGAAATTTATGACAGAGAGTTTAGGTCGAGAGATATTAAGTACAGCCAC 1200
 1237 AAGAAAGAAATTTATGACAGAGAGTTTAGGTCGAGAGATATTAAGTACAGCCAC 1296
 1201 TACATGAAGAGATGATTCCTCCCAATGATATTCATGCTCAAGTTTAAAGCCAGTG 1260
 1297 TACATGAAGAGATGATTCCTCCCAATGATATTCATGCTCAAGTTTAAAGCCAGTG 1356
 1261 GATGTCACCTGCTCTAGATCAAAATACGTGAAGACTGTGTGCTTGCCTGATGGTCC 1320
 1357 GATGTCACCTGCTCTAGATCAAAATACGTGAAGACTGTGTGCTTGCCTGATGGTCC 1416
 1321 TTTCCCTCTGGAGTGAGTGCACATCTGCTGGTGGGTGTTACAGAAACAGAAAGGG 1380
 1417 TTTCCCTCTGGAGTGAGTGCACATCTGCTGGTGGGTGTTACAGAAACAGAAAGGG 1476
 1381 TCCCGCCAGCTCCTGGATGCGCAAAAGTCAAGCTGATTCGCAACACTTTGTGCAACTCCCGC 1440
 1477 TCCCGCCAGCTCCTGGATGCGCAAAAGTCAAGCTGATTCGCAACACTTTGTGCAACTCCCGC 1536
 1441 CAACCTATGACACATGATGATGACAGATGATGCTGCGGAAATCTCAGAAACCT 1500
 1537 CAACCTATGACACATGATGATGACAGATGATGCTGCGGAAATCTCAGAAACCT 1596
 1501 GGGCAAGACACTGCGAGGGTGACTCTGGAGGCCCCCTGACCTGTGAGAGAGGAGCGGAC 1560
 1597 GGGCAAGACACTGCGAGGGTGACTCTGGAGGCCCCCTGACCTGTGAGAGAGGAGCGGAC 1656
 1561 TACTAGCTCTATGGGATGATGATGCTGGGCTGAGAGTGTGAGAGAGGCGGAGGCTTAC 1620

Db 1657 TACTACGCTATGGATAGTGAAGTGGGCTGGAGTGGGAGAGGCGAGGGTCTAC 1716
 Qy 1621 ACCCAAGTTACCAATTCCTGAATTCGATCAAAAGCCACCATCAAAAGTGAAGTGGCTTC 1680
 Db 1717 ACCCAAGTTACCAATTCCTGAATTCGATCAAAAGCCACCATCAAAAGTGAAGTGGCTTC 1776
 Qy 1681 TAA 1683
 Db 1777 TAA 1779
 RESULT 5
 AAC76693
 ID AAC76693 standard; cDNA; 3623 BP.
 XX
 AC AAC76693;
 XX
 DT 08-FEB-2001 (first entry)
 XX
 DE Human ORFX ORF2248 polynucleotide sequence SEQ ID NO:4495.
 XX
 KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
 KW vulnerable; antiproliferative; antiparkinsonian; nontropic; neuroprotective;
 KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
 KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
 KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
 KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
 KW cardiovascular disease; diabetes mellitus; erythematoidism; SCID; AIDS;
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
 KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
 KW thrombosis; contraceptive; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200058473-A2.
 XX
 PD 05-OCT-2000.
 XX
 PF 31-MAR-2000; 2000WO-US008621.
 XX
 PR 31-MAR-1999; 98US-0127607P.
 PR 02-APR-1999; 98US-0127636P.
 PR 05-APR-1999; 98US-0127728P.
 PR 30-MAR-2000; 2000US-00540763.
 XX
 PA (CURA-) CURAGEN CORP.
 XX
 PI Shinketsu RA, Leach M;
 XX
 DR WPI; 2000-602362/57.
 XX
 DR P-PSDB; AAB42484.
 XX
 DR Novel nucleic acids and peptides derived from open reading frame X,
 PT useful for treating e.g. cancers, proliferative disorders,
 PT neurodegenerative disorders and cardiovascular disease.
 XX
 PS Claim 5; Page 3679-3681; 5507pp; English.
 XX
 CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
 CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
 CC sequences have activities such as: cytostatic; hepatotropic; vulnerable;
 CC antiproliferative; antiparkinsonian; nontropic; neuroprotective; osteopathic;
 CC anticonvulsant; antirheumatic; immunosuppressant; immunostimulant;
 CC cardiant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive;
 CC dermatological; immunosuppressive; antiinflammatory; antibacterial;
 CC antiviral; antifungal; antirheumatic; antithyroid; and antianemic. The
 CC sequences can be used for determining the presence of or predisposition
 CC to, or preventing or treating pathological conditions associated with an
 CC ORFX-associated disorder. The nucleic acids can be used to express ORFX.

CC proteins in gene therapy vectors. The proteins and nucleic acids may be
 CC used to treat cancers, proliferative disorders, neurodegenerative
 CC disorders, osteoarthritis, graft vs host disease, cardiovascular disease,
 CC diabetes mellitus, hypertension, hypothyroidism, cholesterol ester
 CC storage, systemic lupus erythematosus, severe combined immunodeficiency
 CC (SCID), AIDS, viral bacterial or fungal infection, malaria, autoimmune
 CC disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and
 CC cartilage damage, nocturnal haemoglobinuria, anti-inflammatory disease; to
 CC enhance coagulation; to inhibit thrombosis; and as a contraceptive
 XX

SQ Sequence 3623 BP; 848 A; 1005 C; 930 G; 840 T; 0 U; 0 Other;
 Query Match 99.5%; Score 1675; DB 3; Length 3623;
 Best Local Similarity 99.7%; Pred. No. 0;
 Matches 1678; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 ATGTTTGCAGGATCTGATCTCCATGTTCTGCTTTAATGGCTCTGGTGGGAAAGACA 60
 DB 97 ATGTTTGCAGGATCTGATCTCCATGTTCTGCTTTAATGGCTCTGGTGGGAAAGACA 156
 QY 61 GCCTGTGGTCTCCCTGATCTCTTATTGAAAGCCTGGACCCAGACTGACCCCTGAC 120
 DB 157 GCCTGTGGTCTCCCTGATGCTTTATTGAAAGCCTGGACCCAGACTGACCCCTGAC 216
 QY 121 CAGTATGATTACAGCTACGAGATTATTAATCAGGAAGAGAACACAGTAGCACACTTACC 180
 DB 217 CAGTATGATTACAGCTACGAGATTATTAATCAGGAAGAGAACACAGTAGCACACTTACC 276
 QY 181 CATGCTGAGATCTGATCTGATCTACACTGAGGACCAAGCTGATCCATGCGAGCCCAAC 240
 DB 277 CACGCTGAGATCTGATCTGATCTACACTGAGGACCAAGCTGATCCATGCGAGCCCAAC 336
 QY 241 CCCTGTGAACACGGTGGGAGCTGCTCTCATGGGAGCACCTTACATGAGCTGCTG 300
 DB 337 CCCTGTGAACACGGTGGGAGCTGCTCTCATGGGAGCACCTTACATGAGCTGCTG 396
 QY 301 GCTCTTCTCTGGGAAATAGTGTGAGAAAGTGCAAAATAGCTGCAAGGACAAACCCATGT 360
 DB 397 GCTCTTCTCTGGGAAATAGTGTGAGAAAGTGCAAAATAGCTGCAAGGACAAACCCATGT 456
 QY 361 GSCCGGGGCAATGCTCATACCAGAGTCTCTCTACTACCGCTGCTGTGTAACAC 420
 DB 457 GSCCGGGGCAATGCTCATACCAGAGTCTCTCTACTACCGCTGCTGTGTAACAC 516
 QY 421 CTTTACAGAGTCCAGCTGCTCCCAAGTGTCTCTGATGAGCCAAAACCCCTGCGAG 480
 DB 517 CTTTACAGAGTCCAGCTGCTCCCAAGTGTCTCTGATGAGCCAAAACCCCTGCGAG 576
 QY 481 AATGGGGTCACTGCTCCCGGCAATAGCGAGATCCAAAGTTCACTGCTGCTGCTCCGAC 540
 DB 577 AATGGGGTCACTGCTCCCGGCAATAGCGAGATCCAAAGTTCACTGCTGCTGCTCCGAC 636
 QY 541 CAGTTCAAGGGGAAATCTGTGAATAGGTTCTGATGACTCTATGTTGGCGATGGCTAC 600
 DB 637 CAGTTCAAGGGGAAATCTGTGAATAGGTTCTGATGACTCTATGTTGGCGATGGCTAC 696
 QY 601 TCTTACCGGAAATGATAGGACAGTCAACAGATGGTGGCTTTACTGGAATCC 660
 DB 697 TCTTACCGGAAATGATAGGACAGTCAACAGATGGTGGCTTTACTGGAATCC 756
 QY 661 CACCTCCTCTTGCAGGAAATTTACAACATGTTTATGGAGGATGCTGAAACCCATGGGATT 720
 DB 757 CACCTCCTCTTGCAGGAAATTTACAACATGTTTATGGAGGATGCTGAAACCCATGGGATT 816
 QY 721 GGGGAACACAAATTTCTGAGAAACCCAGATGGGACGAAAGCCCTGTGCTTTATTA 780
 DB 817 GGGGAACACAAATTTCTGAGAAACCCAGATGGGACGAAAGCCCTGTGCTTTATTA 876
 QY 781 GTTACCAATGACAGGTGAAATGGGAATACTGTGATGCTCAGCTGCTCAGCCGAGGAC 840
 DB 877 GTTACCAATGACAGGTGAAATGGGAATACTGTGATGCTCAGCTGCTCAGCCGAGGAC 936
 QY 841 GTTGCTTACCCAGAGAAAGCCCACTGAGGCCATCAACCAAGCTTCCGGGGTTTGACTCC 900

DB 937 GTTGCTTACCCAGAGAAAGCCCACTGAGCCATCAACCAAGCTTCCGGGGTTGACTCC 996
 QY 901 TGTGAAAGACTGAGATACAGAGAGAGAGATCAAGAGAAATCTATGGAGCTTTAAGAGC 960
 DB 997 TGTGAAAGACTGAGATACAGAGAGAGAGATCAAGAGAAATCTATGGAGCTTTAAGAGC 1056
 QY 961 ACGGGGGCAAGCACCATTGGAGGGTCCCTCCAGTCTCCCTGCTGCTGACCATCTCC 1020
 DB 1057 ACGGGGGCAAGCACCATTGGAGGGTCCCTCCAGTCTCCCTGCTGCTGACCATCTCC 1116
 QY 1021 ATGCCCCAGGGCCACTTCTGTGTGGGGGCTGATCCACCCCTGCTGGGTGCTCACTGCT 1080
 DB 1117 ATGCCCCAGGGCCACTTCTGTGTGGGGCACTGATCCACCCCTGCTGGGTGCTCACTGCT 1176
 QY 1081 GCCCACTGCACCGACATAAATAAACACAGACATCTAAAGGTGGTCTAGGGGACAGGACCTG 1140
 DB 1177 GCCCACTGCACCGACATAAATAAACACAGACATCTAAAGGTGGTCTAGGGGACAGGACCTG 1236
 QY 1141 AAGAAAGAGAAATTTTCATGAGCAGAGCTTTTGGGTGCAAGATATTCAGGTACAGCCAC 1200
 DB 1237 AAGAAAGAGAAATTTTCATGAGCAGAGCTTTTGGGTGCAAGATATTCAGGTACAGCCAC 1296
 QY 1201 TACAATGAAGAGATGAGATTCCCAATGATATTTGCAATTTGCTCAAGTTAAAGCCAGTG 1260
 DB 1297 TACAATGAAGAGATGAGATTCCCAATGATATTTGCAATTTGCTCAAGTTAAAGCCAGTG 1356
 QY 1261 GATGTCATCTGCTCTAGAAATCCAAATACGTGAAGACTGCTGCTGCTGCTGATGGTCC 1320
 DB 1357 GATGTCATCTGCTCTAGAAATCCAAATACGTGAAGACTGCTGCTGCTGCTGATGGTCC 1416
 QY 1321 TTTCCCTCTGGGAGTGAAGTCCACATCTCTGGCTGGGGTGTACAGAAACAGGAAAGGG 1380
 DB 1417 TTTCCCTCTGGGAGTGAAGTCCACATCTCTGGCTGGGGTGTACAGAAACAGGAAAGGG 1476
 QY 1381 TCCCGCAGCTCTGATGCTCAAGCTGATTCGCAACACTTTGTCGAACTCCCGC 1440
 DB 1477 TCCCGCAGCTCTGATGCTCAAGCTGATTCGCAACACTTTGTCGAACTCCCGC 1536
 QY 1441 CAACTCTATGACCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1500
 DB 1537 CAACTCTATGACCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1596
 QY 1501 GSCCAGACACTGCTCCAGGGTGAATCTGAGGCCCCCTGACCTGTGAGAGAGAGCCGAC 1560
 DB 1597 GSCCAGACACTGCTCCAGGGTGAATCTGAGGCCCCCTGACCTGTGAGAGAGAGCCGAC 1656
 QY 1561 TACTAGCTCTATGGGATAGTGAAGTGGGCTGGAGTGTGAGAAAGGAGGAGGAGGCTTAC 1620
 DB 1657 TACTAGCTCTATGGGATAGTGAAGTGGGCTGGAGTGTGAGAAAGGAGGAGGAGGCTTAC 1716
 QY 1621 ACCCAAGTTACAAATTTCTGAAATGGATCAAGCCACCATCAAAAGTGAAGTGGCTTC 1680
 DB 1717 ACCCAAGTTACAAATTTCTGAAATGGATCAAGCCACCATCAAAAGTGAAGTGGCTTC 1776
 QY 1681 TAA 1683
 DB 1777 TAA 1779

RESULT 6
 AAA44763
 ID AAA44763 standard; cdna; 617 BP.
 XX
 AC AAA44763;
 XX
 DT 21-AUG-2000 (first entry)
 XX
 DE Human secreted expressed sequence tag SEQ ID NO:1338.
 XX
 KW Human; mouse; chicken; rat; secreted expressed sequence tag; EST;
 KW expressed sequence tag; EST; probe; chemotactic; proliferative;
 KW immunomodulatory; haematopoietic; chemokinetic; analgesic; haemostatic;

thrombolytic; antiinflammatory; cytostatic; antibacterial; antifungal;
KW antiviral; antidiabetic; antiasthmatic; vulnary; antiparkinsonian;
KW antitumor; osteopathic; neuroprotective; nootropic; antipsoriatic;
KW cerebroprotective; anticonvulsant; antidepressant; gene therapy; vaccine;
KW autoimmune disorder; multiple sclerosis; allergic condition; ulcer;
KW insulin dependent diabetes; burn; osteoporosis; osteoarthritis;
KW lymphoid cell deficiency; asthma; myeloid cell deficiency; stroke;
KW central nervous system disorder; Alzheimer's disease; stroke;
KW Parkinson's disease; Huntington's disease; coagulation disorder;
KW haemophilia; thrombosis; inflammatory disorder; Crohn's disease; tumour;
KW infection; depression; psoriasis; ss.
XX Homo sapiens.
OS

XX WO200021991-A1.
XX 20-APR-2000.
XX 15-OCT-1999; 99WO-US024206.
XX 15-OCT-1998; 98US-0104436P.
XX (GEMY) GENETICS INST INC.
XX Jacobs K, McCoy JM, Lavallie ER, Collins-Racie LA, Evans C;
PI Merberg D, Treacy M, Bowman MR;
PI
XX WPI; 2000-317938/27.
XX

XX Isolated polynucleotides, and encoded proteins, comprising secreted
PT expressed sequence tags (SESTs), useful for treating various disorders
PT such as autoimmune, infectious, and central nervous system disorders.
XX
XX Claim 1; Page 549; 803pp; English.
XX

XX AAA43426 to AAA45925 represent specifically claimed expressed
CC sequence tags (SESTs), isolated from human, mouse, chicken and rat tissue
CC sources. The SESTs can have a range of activities depending on the
CC tissues they were isolated from. The activities include: chemotactic;
CC proliferative; immunomodulatory; haematopoietic; chemokinetic; analgesic;
CC haemostatic; thrombolytic; antiinflammatory; cytostatic; antibacterial;
CC antifungal; antiviral; antidiabetic; antiasthmatic; vulnary; antitumor;
CC osteopathic; neuroprotective; nootropic; antiparkinsonian; antipsoriatic;
CC cerebroprotective; anticonvulsant; and antidepressant. The SESTs can be
CC used for gene therapy and in vaccines. The SESTs are useful as probes for
CC the identification and isolation of full-length cDNAs and genomic DNA
CC molecules which correspond to the SESTs. Proteins encoded by the SESTs
CC are useful in assays for determining biological activity and raising
CC antibodies. They may be useful for treatment of autoimmune disorders
CC (multiple sclerosis, insulin dependent diabetes), allergic conditions
CC (asthma), myeloid or lymphoid cell deficiencies, wounds, burns, ulcers,
CC osteoporosis, osteoarthritis, central nervous system disorders
CC (Alzheimer's, Parkinson's, Huntington's disease, stroke), coagulation
CC disorders (haemophilia, thrombosis), inflammatory disorders (Crohn's
CC disease), tumours, bacterial, fungal or viral infections, depression and
CC psoriasis. AAA45926 to AAA45931 represent linker variants which are given
CC in the exemplification of the present invention
XX

SQ Sequence 617 BP; 165 A; 169 C; 148 G; 135 T; 0 U; 0 Other;
Query Match 31.3%; Score 526.4; DB 3; Length 617;
Best Local Similarity 99.8%; Pred. No. 1.5e-139;
Matches 527; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ATGTTTGCAGGATGTCGATCTCCATGTCCTGCTGTTATGCTGTGGGAAAGACA 60
Db 84 ATGTTTGCAGGATGTCGATCTCCATGTCCTGCTGTTATGCTGTGGGAAAGACA 143
Qy 61 GCCTGTGGGTTCTCCCTGATGTCCTTTATTGGAAGCCTGGACCCAGACTGGACCCCTGAC 120
Db 144 GCCTGTGGGTTCTCCCTGATGTCCTTTATTGGAAGCCTGGACCCAGACTGGACCCCTGAC 203
Qy 121 CAGTATGATTACAGTACGAGGATTAATATCAGGAGAGAACCCAGTACACACTTACC 180

Db 204 CAGTATGATTACAGTACGAGGATTAATATCAGGAGAGAACCCAGTACACACTTACC 263
Qy 181 CATGCTCAGAAATCCTGACTGGTACTACACTGAGGACCAAGCTGATCCATGCCGCCAAC 240
Db 264 CAGCTGAGAAATCCTGACTGGTACTACACTGAGGACCAAGCTGATCCATGCCGCCAAC 323
Qy 241 CCCTGTGAAACCGTGGGAGTCCCTGCTGCTCCATGAGGAGACCTTCACATGACGCTGCTG 300
Db 324 CCCTGTGAAACCGTGGGAGTCCCTGCTGCTCCATGAGGAGACCTTCACATGACGCTGCTG 383
Qy 301 GCTCTTTCTCTGGGAAATAAGTGTGAGAAATACGTCGAGGACCAAGCTGATCCATGCCGCCAAC 360
Db 384 GCTCTTTCTCTGGGAAATAAGTGTGAGAAATACGTCGAGGACCAAGCTGATCCATGCCGCCAAC 443
Qy 361 GCGCGGGGCAATGCTTCATTAATCCAGAGTCTCCCTACTACCGCTGTGTCTGTAAACAC 420
Db 444 GCGCGGGGCAATGCTTCATTAATCCAGAGTCTCCCTACTACCGCTGTGTCTGTAAACAC 503
Qy 421 CCTTACACAGTCCCGAGTCTCCCAAGTGTCTCTGATGAGGACCAAGCTGATCCATGCCGCCAAC 480
Db 504 CCTTACACAGTCCCGAGTCTCCCAAGTGTCTCTGATGAGGACCAAGCTGATCCATGCCGCCAAC 563
Qy 481 AATGGGCTACCTGCTCCCGGATTAAGCGGAGATCCAGTTCACCTGT 528
Db 564 AATGGGCTACCTGCTCCCGGATTAAGCGGAGATCCAGTTCACCTGT 611

RESULT 7
ACH21030
ID ACH21030 standard; cDNA; 428 BP.
XX
AC ACH21030;
XX
DT 13-OCT-2003 (first entry)
XX
DE Human adult liver cDNA #642.
XX
KW Human; ss; sequencing by hybridisation; SH; expressed sequence tag; EST;
KW Genome mapping; biodiversity; genetic disorder.
XX
OS Homo sapiens.
XX
PN US2003073623-A1.
PD 17-APR-2003.
XX
PF 30-JUL-2001; 2001US-00918995.
XX
PR 30-JUL-2001; 2001US-00918995.
XX
PA (DRMA/) DRMANAC R T.
PA (LABA/) LABAT I.
PA (STAC/) STACHE-CRAIN B.
PA (DICK/) DICKSON M C.
PA (JONE/) JONES L W.
XX
PI Drmanac RT, Labat I, Stache-Crain B, Dickson MC, Jones LW;
XX
XX WPI; 2003-615964/58.

XX New polynucleotide sequences obtained from various cDNA libraries, useful
PT as hybridization probes, as oligomers for PCR, for chromosome and gene
PT mapping, in the recombinant production of protein, or in generating
PT antisense DNA or RNA.
XX
PS Claim 1; SEQ ID NO 8242; 44pp; English.
XX
CC The invention relates to an isolated polynucleotide comprising any one of
CC 38043 cDNA sequences, appearing as ACH12789-ACH50831, whose sequence was
CC determined by the technique of SH (sequencing by hybridisation). Also
CC included is a purified polypeptide comprising a sequence corresponding to
CC a reading frame of the novel polynucleotide. The nucleic acid sequences

CC are useful in diagnostics as expressed sequence tags (EST) for
CC identifying expressed genes or for physical mapping of the human genome,
CC in forensics, in assessing biodiversity, or in identifying mutations
CC responsible for genetic disorders and other traits. The nucleotide
CC sequences are also useful as hybridisation probes, as oligomers for PCR,
CC for chromosome and gene mapping, in the recombinant production of
CC protein, or in generating antisense DNA or RNA. The purified polypeptide
CC is useful for generating antibodies specific for it. The present sequence
CC is one of the 38043 isolated cDNA/EST sequences. Note: The sequence data
CC for this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from USPTO at
CC seqdata.uspto.gov/sequence.html?DocID=20030073623
XX
SQ Sequence 428 BP; 119 A; 108 C; 105 G; 96 T; 0 U; 0 Other;

Query Match 21.9%; Score 368.4; DB 8; Length 428;
Best Local Similarity 99.7%; Pred. No. 1.6e-94;
Matches 369; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 ATGTTTCCAGGATGCTGATCTCCATGTTCTGCTGTTAATGCTCTGTGGGAAAGACA 60
DB 59 ATGTTTCCAGGATGCTGATCTCCATGTTCTGCTGTTAATGCTCTGTGGGAAAGACA 118
QY 61 GCCTGTGGTTCCTCCGTGCTCTTTATGGAAGCCTGGACCCAGACTGGACCCCTGAC 120
DB 119 GCCTGTGGTTCCTCCGTGCTCTTTATGGAAGCCTGGACCCAGACTGGACCCCTGAC 178
QY 121 CAGTATGATTACACTACGAGGATTATAATCAGGAAGACACACAGTACACACTTACC 180
DB 179 CAGTATGATTACACTACGAGGATTATAATCAGGAAGACACACAGTACACACTTACC 238
QY 181 CATGCTGAGAAATCCTGACTGTGTACTACACTGAGACCAAGCTGATCCATCCAGCCCAAC 240
DB 239 CATGCTGAGAAATCCTGACTGTGTACTACACTGAGACCAAGCTGATCCATCCAGCCCAAC 298
QY 241 CCCTGTGAACACGCTGGGACTGCTCGTCCATGGGAGCACCTTACATGACGTGCTG 300
DB 299 CCCTGTGAACACGCTGGGACTGCTCGTCCATGGGAGCACCTTACATGACGTGCTG 358
QY 301 GCTCTTTCTCTGGGAATAGTGTGAGAAAGTCAAAATACGTGCAAGGCAACCCATGT 360
DB 359 GCTCTTTCTCTGGGAATAGTGTGAGAAAGTCAAAATACGTGCAAGGCAACCCATGT 418
QY 361 GGCCGGGGCC 370
DB 419 GGCCGGGGCC 428

RESULT 8
ABI99281
ID ABI99281 standard; cDNA; 397 BP.
XX
AC ABI99281;
XX
DT 07-MAR-2002 (first entry)
XX
DE Mouse ischaemic condition related cDNA sequence SEQ ID NO:123.
XX
KW Mouse; ischaemia; compressive ischaemia; occlusive ischaemia;
XX vasospastic ischaemia; ischaemic condition; ischaemic disease; ss.
XX
OS Mus musculus.
XX
PN WO200188188-A2.
XX
XX 22-NOV-2001.
XX
XX 18-MAY-2001; 2001WO-JP004192.
XX
XX 18-MAY-2000; 2000JP-00145977.
XX
XX (UYNI-) UNIV NIHON SCHOOL JURIDICAL PERSON.
XX

PI Ishikawa K, Asai S, Takahashi Y, Nagata T, Ishii Y;
XX WPI; 2002-034733/04.
XX
XX Examining the ischemic condition (e.g. occlusive ischemia) by measuring
XX expression levels of particular genes defined in the specification or by
XX determining the expression profile of a gene group comprising these
XX genes.
XX
XX Claim 2; Page 351-352; 2690pp; English.
XX
XX The present invention describes a method for examining ischaemic
XX conditions, comprising measuring the expression levels of particular
XX genes (I) in a test sample or determining the expression profile of a
XX gene group in the sample comprising genes selected from (I). The method
XX is useful for examining the ischaemic condition (e.g. compressive
XX ischaemia, occlusive ischaemia or vasospastic ischaemia) by measuring
XX expression levels of particular genes (ABI99202 to ABI99912, encoding the
XX protein sequences in AB57020 to AB57374) or by determining the
XX expression profile of a gene group comprising these genes. The expression
XX levels or expression profiles produced by these genes are used as an
XX indicator when screening for ischaemic condition-improving drugs or
XX therapeutics for ischaemic diseases. ABI99913 and ABI99914 represent PCR
XX primers for a mouse ischaemic condition related sequence, which are used
XX in the exemplification of the present invention
XX
XX Sequence 397 BP; 103 A; 105 C; 111 G; 78 T; 0 U; 0 Other;

Query Match 14.3%; Score 241; DB 6; Length 397;
Best Local Similarity 84.4%; Pred. No. 3.5e-58;
Matches 271; Conservative 0; Mismatches 50; Indels 0; Gaps 0;
QY 1363 ACAGAAACAGGAAAGGTCCTCCGCGAGCTCTGATGCCAAGTCAAGCTGATGCCAAC 1422
DB 13 ACAGAAACAGGAAAGGTCCTCCGCGAGCTCTGATGCCAAGTCAAGCTGATGCCAAC 72
QY 1423 ACTTTGTGCAACTCCCGCAACTCTATGACACCATGATGATGATGATGATGATG 1482
DB 73 CTTTGTGCAACTCCCGCAACTCTATGACACCATGATGATGATGATGATGATG 132
QY 1483 GGAATCTTCAGAACCTGGGCAACACCTGCCAGGTGACTCTGGAGGCCCTGACC 1542
DB 133 GGGAACTCTCAGAACCCCGGATCAGACACCTGCCAGGTGACTCTGGGGGGCCCTTAACC 192
QY 1543 TGTGAGAGGACGGCACCTTACTTACCTCTATGGATAGTACCTGGGGCTGAGTGTGAG 1602
DB 193 TGTGAGAGGATGGAACTTACTTACTTACCTTACCTGAGGATGTAAGCTGGGCCAGGAATGTGGG 252
QY 1603 AAGAGCCAGGGGTCTACACCCCAAGTTACCAATTCCTGAATTGGATCAAGCCACCATC 1662
DB 253 AAGAGCCAGGGGTCTACACTCAAGTCAAGTCAAGTCTCTGAATTGGATAAGACCACCATG 312
QY 1663 AAAAGTGAAAGTGGCTTCTAA 1683
DB 313 CACAGGAGGCTGGCCTCTGA 333

RESULT 9
ABA58323/c
ID ABA58323 standard; DNA; 451 BP.
XX
AC ABA58323;
XX
XX 01-FEB-2002 (first entry)
XX
XX Human foetal liver single exon nucleic acid probe #6628.
XX
XX Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
XX
XX Homo sapiens.
XX
XX WO200157277-A2.
XX


```
PD 09-AUG-2001.
XX 30-JAN-2001; 2001WO-US000669.
XX 04-FEB-2000; 2000US-0180312P.
XX 26-MAY-2000; 2000US-0207456P.
XX 30-JUN-2000; 2000US-00608408.
XX 03-AUG-2000; 2000US-00632366.
XX 21-SEP-2000; 2000US-0234687P.
XX 27-SEP-2000; 2000US-0236359P.
XX 04-OCT-2000; 2000GB-00024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-483447/52.
XX Human genome-derived single exon nucleic acid probes useful for analyzing
XX gene expression in human fetal liver.
XX Claim 1; SEQ ID NO 628; 639pp + Sequence Listing; English.
XX The invention relates to a single exon nucleic acid probe for measuring
XX human gene expression in a sample derived from human foetal liver. The
XX single exon nucleic acid probes may be used for predicting, measuring and
XX displaying gene expression in samples derived from human fetal liver. The
XX present sequence is a single exon nucleic acid probe of the invention.
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences
XX Sequence 451 BP; 116 A; 102 C; 127 G; 106 T; 0 U; 0 Other;
XX Query Match 7.1%; Score 120.2; DB 4; Length 451;
XX Best Local Similarity 97.6%; Pred. No. 1.1e-23;
XX Matches 122; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 324 TCAGAAAGTGCAGAAATACGTGCAAGGACCAACCCATGTGCGCGGGGCCAATGTCTCATTAC 383
Db 451 TCCTACAGTGCAGAAATACGTGCAAGGACCAACCCATGTGCGCGGGGCCAATGTCTCATTAC 392
QY 384 CCAGAGTCCCTCCTACTACCGCTGTCTGTGTAACACCCCTTACACAGGTCCCAGCTGCTC 443
Db 391 CCAGAGTCCCTCCTACTACCGCTGTCTGTGTAACACCCCTTACACAGGTCCCAGCTGCTC 332
QY 444 CCAAG 448
Db 331 CCAAG 327
RESULT 10
AAI37947/c
ID AAI37947 standard; DNA; 451 BP.
XX AC AAI37947;
XX DT 17-OCT-2001 (first entry)
XX DE Probe #6633 used to measure gene expression in human placenta sample.
XX KW Probe; microarray; human; placenta; antenatal diagnosis;
XX KW Genetic disorder; ss.
XX OS Homo sapiens.
XX PN WO200157272-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US000663.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
```

```
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488897/53.
XX Human genome-derived single exon nucleic acid probes useful for analyzing
XX gene expression in human placenta.
XX Claim 25; SEQ ID NO 6633; 654pp; English.
XX The present invention relates to single exon nucleic acid probes (SENP).
XX The present sequence is one such probe. The probes are useful for
XX producing a microarray for predicting, measuring and displaying gene
XX expression in samples derived from human placenta. The probes are useful
XX for antenatal diagnosis of human genetic disorders
XX Sequence 451 BP; 116 A; 102 C; 127 G; 106 T; 0 U; 0 Other;
XX Query Match 7.1%; Score 120.2; DB 4; Length 451;
XX Best Local Similarity 97.6%; Pred. No. 1.1e-23;
XX Matches 122; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 324 TCAGAAAGTGCAGAAATACGTGCAAGGACCAACCCATGTGCGCGGGGCCAATGTCTCATTAC 383
Db 451 TCCTACAGTGCAGAAATACGTGCAAGGACCAACCCATGTGCGCGGGGCCAATGTCTCATTAC 392
QY 384 CCAGAGTCCCTCCTACTACCGCTGTCTGTGTAACACCCCTTACACAGGTCCCAGCTGCTC 443
Db 391 CCAGAGTCCCTCCTACTACCGCTGTCTGTGTAACACCCCTTACACAGGTCCCAGCTGCTC 332
QY 444 CCAAG 448
Db 331 CCAAG 327
RESULT 11
AAK32096/c
ID AAK32096 standard; DNA; 451 BP.
XX AC AAK32096;
XX DT 06-NOV-2001 (first entry)
XX DE Human bone marrow expressed single exon probe SEQ ID NO: 6653.
XX KW Human; bone marrow expressed exon; gene expression analysis; probe;
XX KW microarray; cancer; leukaemia; lymphoma; myeloma; ss.
XX OS Homo sapiens.
XX PN WO200157276-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US000668.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
```


CC measuring human gene expression in a sample derived from human adult
 CC liver, comprising one of 13109 defined nucleotide sequences given in the
 CC specification (or complements/ fragments). The probe hybridises at high
 CC stringency to a nucleic acid molecule expressed in the human adult liver.
 CC (i) may be used for predicting, measuring and displaying gene expression
 CC in samples derived from human adult liver. The genes identified may be
 CC involved in genetic liver diseases such as cirrhosis,
 CC hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which is
 CC associated with coronary heart disease. AB25011-AB51005 represent human
 CC liver single exon nucleic acid probes of the invention. Note: The
 CC sequence information for this patent does not appear in the printed
 CC specification but was obtained in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 451 BP; 116 A; 102 C; 127 G; 106 T; 0 U; 0 Other;
 Query Match 7.1%; Score 120.2; DB 4; Length 451;
 Best Local Similarity 97.6%; Pred. No. 1.1e-23;
 Matches 122; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 324 TCAGAAAGTGCAGAAATACGTGCAAGGACACCCATGTGGCGGGCCAAATGTCTCATTTAC 383
 Db 451 TCCTACAGTGCAGAAATACGTGCAAGGACACCCATGTGGCGGGCCAAATGTCTCATTTAC 392
 QY 384 CCAGAGTCTCTCTACTACCGCTGTCTGTAAACACCCCTTACACAGGTCCAGCTGCTC 443
 Db 391 CCAGAGTCTCTCTACTACCGCTGTCTGTAAACACCCCTTACACAGGTCCAGCTGCTC 332
 QY 444 CCAAG 448
 Db 331 CCAAG 327
 RESULT 14
 ABS06863/C
 ID ABS06863 standard; DNA; 451 BP.
 XX
 AC ABS06863;
 XX
 DT 19-AUG-2002 (first entry)
 XX
 DE Human genome-derived single exon probe from lung SEQ ID No 6854.
 XX
 KW Human; ds; single exon probe; asthma; lung cancer; COPD; ILD;
 KW Chronic obstructive pulmonary disease; interstitial lung disease;
 KW familial idiopathic pulmonary fibrosis; neurofibromatosis;
 KW tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;
 KW Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;
 KW pulmonary histiocytosis; lymphangioleiomyomatosis; Karagener syndrome;
 KW pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;
 KW Primary ciliary dyskinesia; pulmonary hypertension;
 KW hyaline membrane disease.
 XX
 OS Homo sapiens.
 XX
 PN WO200186003-A2.
 XX
 PD 15-NOV-2001.
 XX
 PF 30-JAN-2001; 2001WO-US000665.
 XX
 PR 04-FEB-2000; 2000US-0180312P.
 PR 26-MAY-2000; 2000US-0207456P.
 PR 30-JUN-2000; 2000US-00608408.
 PR 03-AUG-2000; 2000US-00632366.
 PR 21-SEP-2000; 2000US-0234687P.
 PR 27-SEP-2000; 2000US-0236359P.
 PR 04-OCT-2000; 2000GB-00024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX

DR WPI; 2002-114183/15.
 XX
 PT Spatially-addressable set of single exon nucleic acid probes, used to
 measure gene expression in human lung samples.
 XX
 PS Claim 1; SEQ ID NO 6854; 634bp; English.
 XX
 CC The invention relates to a spatially-addressable set of single exon
 CC nucleic acid probes for measuring gene expression in a sample derived
 CC from human lung comprising single exon nucleic acid probes having one of
 CC 12614 nucleic acid sequences mentioned in the specification, or their
 CC complements or the 12387 open reading frames derived from the 12614
 CC probes. Also included are a microarray comprising the novel set of probes
 CC; the novel set of probes which hybridise at high stringency to a nucleic
 CC acid expressed in the human lung; measuring gene expression in a sample
 CC derived from human lung, comprising (a) contacting the array with a
 CC collection of detectably labeled nucleic acids derived from human lung
 CC mRNA, and (b) measuring the label detectably bound to each probe of the
 CC array; identifying exons in a eukaryotic genome, comprising (a)
 CC algorithmically predicting at least one exon from genomic sequences of
 CC the eukaryote; and (b) detecting specific hybridisation of detectably
 CC labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,
 CC having a fragment identical to the predicted exon, the probe is included
 CC in the above mentioned microarray; assigning exons to a single gene,
 CC comprising (a) identifying exons from genomic sequence by the method
 CC above and (b) measuring the expression of each of the exons in several
 CC tissues and/or cell types using hybridisation to a single exon
 CC microarrays having a probe with the exon, where a common pattern of
 CC expression of the exons in the tissues and/or cell types indicates that
 CC the exons should be assigned to a single gene; a peptide comprising one
 CC of 12011 sequences, mentioned in the specification, or encoded by the
 CC probes/open reading frames (ORF). The probes are used for gene expression
 CC analysis, and for identifying exons in a gene, particularly using human
 CC lung derived mRNA and for the study of lung diseases such as asthma, lung
 CC cancer, chronic obstructive pulmonary disease (COPD), interstitial lung
 CC disease (ILD), familial idiopathic pulmonary fibrosis, neurofibromatosis,
 CC tuberous sclerosis, Gaucher's disease, Niemann-Pick disease, Hermansky-
 CC Pudlak syndrome, sarcoidosis, pulmonary haemosiderosis, pulmonary
 CC histiocytosis, lymphangioleiomyomatosis, pulmonary alveolar proteinosis,
 CC Karagener syndrome, fibrocystic pulmonary dysplasia, primary ciliary
 CC dyskinesia, pulmonary hypertension and hyaline membrane disease. The
 CC present sequence is a single exon probe of the invention. Note: The
 CC sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 451 BP; 116 A; 102 C; 127 G; 106 T; 0 U; 0 Other;
 Query Match 7.1%; Score 120.2; DB 6; Length 451;
 Best Local Similarity 97.6%; Pred. No. 1.1e-23;
 Matches 122; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 324 TCAGAAAGTGCAGAAATACGTGCAAGGACACCCATGTGGCGGGCCAAATGTCTCATTTAC 383
 Db 451 TCCTACAGTGCAGAAATACGTGCAAGGACACCCATGTGGCGGGCCAAATGTCTCATTTAC 392
 QY 384 CCAGAGTCTCTCTACTACCGCTGTCTGTAAACACCCCTTACACAGGTCCAGCTGCTC 443
 Db 391 CCAGAGTCTCTCTACTACCGCTGTCTGTAAACACCCCTTACACAGGTCCAGCTGCTC 332
 QY 444 CCAAG 448
 Db 331 CCAAG 327
 RESULT 15
 ABA70919/c
 ID ABA70919 standard; DNA; 117 BP.
 XX
 AC ABA70919;
 XX
 DT 01-FEB-2002 (first entry)
 XX

DE Human foetal liver single exon nucleic acid probe #19224.
XX
KW Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
XX
OS Homo sapiens.
XX
FN WO200157277-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US000659.
XX
PR 04-FEB-2000; 2000US-0180312P.
XX
PR 26-MAY-2000; 2000US-0207456P.
XX
PR 30-JUN-2000; 2000US-00608408.
XX
PR 03-AUG-2000; 2000US-00632366.
XX
PR 21-SEP-2000; 2000US-0234687P.
XX
PR 27-SEP-2000; 2000US-0236359P.
XX
PR 04-OCT-2000; 2000GB-00024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-483447/52.
XX
PT Human genome-derived single exon nucleic acid probes useful for analyzing
PT Gene expression in human fetal liver.
XX
PS Claim 4; SEQ ID NO 19224; 639pp + Sequence Listing; English.
XX
CC The invention relates to a single exon nucleic acid probe for measuring
CC human gene expression in a sample derived from human foetal liver. The
CC single exon nucleic acid probes may be used for predicting, measuring and
CC displaying gene expression in samples derived from human fetal liver. The
CC present sequence is a single exon nucleic acid probe of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 117 BP; 24 A; 24 C; 40 G; 29 T; 0 U; 0 Other;

Query Match 7.0%; Score 117; DB 4; Length 117;
Best Local Similarity 100.0%; Pred. No. 4.9e-23;
Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 332 TGC AAAATACGTGCAAGGACCAACCCATGTGGCGGGGCCAATGCTCATTTACCCAGATC 391
DB 117 TGC AAAATACGTGCAAGGACCAACCCATGTGGCGGGGCCAATGCTCATTTACCCAGATC 58

QY 392 CTCCTACTACCGCTGTGCTGTAAACACCCCTTACACAGGTCCCAAGTCTCCCAAG 448
DB 57 CTCCTACTACCGCTGTGCTGTAAACACCCCTTACACAGGTCCCAAGTCTCCCAAG 1

Search completed: May 25, 2004, 06:30:49
Job time : 473 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 25, 2004, 05:42:15 ; Search time 3059 Seconds
(without alignments)
16429.568 Million cell updates/sec

Title: US-09-912-559-2
Perfect score: 1683
Sequence: 1 atgttgccaggatgtctga.....aaagtgaagtggtctctaa 1683

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931030276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: em_estba:*
2: em_esthum:*
3: em_esthum:*
4: em_esthum:*
5: em_esthum:*
6: em_esthum:*
7: em_esthum:*
8: em_esthum:*
9: gb_est1:*
10: gb_est2:*
11: gb_est3:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_fod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES			
Result No.	Score	Query Match Length DB ID	Description
1	848.4	50.4 1042 13	EX325239 BX325239
2	767.4	45.6 1008 13	EX463023 BX463023
3	698.2	41.5 891 12	BI761782 BI761782
4	645.4	38.3 889 13	EX431866 BX431866

576	14	572.8	34.0	576	14	CB156834
663	14	533.8	31.7	663	14	CB162144
736	14	454.2	27.0	736	14	CB594245
592	10	451.6	26.8	592	10	BB032018
802	12	447.2	26.6	802	12	BG972579
574	9	436.2	25.9	574	9	A2217892
751	12	435.4	25.9	751	12	BI148082
570	12	432	25.7	570	12	BM508620
929	10	416.4	24.7	929	10	BF788188
665	14	399.4	23.7	665	14	CF169114
791	12	396.4	23.6	791	12	BI332440
546	9	394.2	23.4	546	9	A2373499
608	9	392.6	23.3	608	9	AV601564
902	10	390.4	23.2	902	10	BF780971
682	12	386	22.9	682	12	BI220028
929	10	384.6	22.9	929	10	BF785781
648	14	379.8	22.6	648	14	CF171046
969	10	372.4	22.1	969	10	BF384535
654	10	370.6	22.0	654	10	AW475402
587	12	370.4	22.0	587	12	BG972681
659	10	365.6	21.7	659	10	AW610902
594	12	363.6	21.6	594	12	BM503097
625	14	363.2	21.6	625	14	CF171125
716	10	347.2	20.6	716	10	BF789705
655	10	335.8	20.0	655	10	BB569555
668	13	335.4	19.9	668	13	BY742633
733	14	332.6	19.8	733	14	CB599445
934	12	330.8	19.7	934	12	BI765113
345	14	327.8	19.5	345	14	T68666
787	14	323	19.2	787	14	CB555374
486	9	317.4	18.9	486	9	AA268125
600	10	311.8	18.5	600	10	BF788541
335	14	302.4	18.0	335	14	R89458
104	14	293.2	17.4	104	14	T77362
610	14	286.2	17.0	610	14	CA944054
521	12	279.8	16.6	521	12	BI221626
912	12	275.8	16.4	912	12	BG972626
366	14	274	16.3	366	14	T93666
434	14	264.4	15.7	434	14	T84369
781	10	258.6	15.4	781	10	BF533788
458	14	256.2	15.2	458	14	R10295

ALIGNMENTS

RESULT 1
BX325239
LOCUS
DEFINITION
BX325239 Homo sapiens T CELLS (JURKAT CELL LINE) COT 10-NORMALIZED
Homo sapiens cDNA clone CS0DJ010YL24 5-PRIME, mRNA sequence.
ACCESSION
BX325239
VERSION
BX325239.1
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 1042)
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished (2001)
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 5634.f For
more information about this cluster, see
http://www.genoscope.cns.fr/cgi-bin/cluster.cgi?seq=CS0AJ010DF12QP1
cluster=5634.f. Contact : Feng Liang Email : fliang@lifetech.com
URL : http://fulllength.invitrogen.com/ invitrogen Corporation 1600
Paradise Avenue Genoscope sequence ID : CS0AJ010DF12QP1.

```
FEATURES
  source      Location/Qualifiers
1..1042
  /organism="Homo sapiens"
  /mol_type="mRNA"
  /db_xref="taxon:9606"
  /clone="CS0101Y124"
  /cell_type="T CELLS (JURKAT CELL LINE) COT 10-NORMALIZED"
  /cell_line="JURKAT"
  /clone_lib="Homo sapiens T CELLS (JURKAT CELL LINE) COT 10-NORMALIZED"
  10-NORMALIZED
  /notes="1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN
Query Match      50.4%; Score 848.4; DB 13; Length 1042;
Best Local Similarity 98.1%; Pred. No. 4.9e-231;
Matches 883; Conservative 6; Mismatches 7; Indels 4; Gaps 3;

QY 1 ATGTTTGCCAGGATGCTGATCTCCATGTTCTGCTTAAATGGCTCTGGTGGGAAAGACA 60
DB 146 ATGTTTGCCAGGATGCTGATCTCCATGTTCTGCTTAAATGGCTCTGGTGGGAAAGACA 205
QY 61 GCTGTGGGTTCTCCCTGATGCTTTATTGGAAGCTCGAGCCAGACTGGACCCCTGAC 120
DB 206 GCCTGTGGGTTCTCCCTGATGCTTTATTGGAAGCTCGAGCCAGACTGGACCCCTGAC 265
QY 121 CAGTATGATTACAGCTACGAGGATTAATATCAGGAAGAGAACACCAAGTAGCACATTAC 180
DB 266 CAGTATGATTACAGCTACGAGGATTAATATCAGGAAGAGAACACCAAGTAGCACATTAC 325
QY 181 CATGCTGAGGAATCCTGACTGCTACTACATGAGGACCAAGCTGATCCATGCCAGGCCAAC 240
DB 326 CATGCTGAGGAATCCTGACTGCTACTACATGAGGACCAAGCTGATCCATGCCAGGCCAAC 385
QY 241 CCCTGTGAACACGGTGGGACTGCTGCTCCATGGGAGCACCTTCACATGCGAGCTGCTG 300
DB 386 CCCTGTGAACACGGTGGGACTGCTGCTCCATGGGAGCACCTTCACATGCGAGCTGCTG 445
QY 301 GCTCCTTTCTCTGGGAATAAGTGTGAGAAAGTGCAGAAATACGTCGCAAGCAACCCATGT 360
DB 446 GCTCCTTTCTCTGGGAATAAGTGTGAGAAAGTGCAGAAATACGTCGCAAGCAACCCATGT 505
QY 361 GCGCGGGGCAATGCTCANTACCCAGAGTCCTCCCTACTACCGCTGTGCTGTGTAACAC 420
DB 506 GCGCGGGGCAATGCTCANTACCCAGAGTCCTCCCTACTACCGCTGTGCTGTGTAACAC 565
QY 421 CATTACACAGTCCCGACTGCTCCCAAGTGGTTCCTGTATGTCAGGCAACACCCCTGCCAG 480
DB 566 CATTACACAGTCCCGACTGCTCCCAAGTGGTTCCTGTATGTCAGGCAACACCCCTGCCAG 625
QY 481 AATGGGCTACTGCTCCGGCATACGAGGATCCAGTTCCAGTTCAGCTGCTGCTGCCGAC 540
DB 626 AATGGGCTACTGCTCCGGCATACGAGGATCCAGTTCAGTTCAGCTGCTGCTGCCGAC 685
QY 541 CAGTTCAAGGGGAAATCTGTGAATAGGTTCTGATGACTGCTATGTTGGCGATGGCTAC 600
DB 586 CAGTTCAAGGGGAAATCTGTGAATAGGTTCTGATGACTGCTATGTTGGCGATGGCTAC 745
QY 601 TCTTACCGAGGAAATGAATAGACAGTCAACAGCATGCTGCTCTTTACTGGAAGTCC 660
DB 746 TCTTACCGAGGAAATGAATAGACAGTCAACAGCATGCTGCTCTTTACTGGAAGTCC 805
QY 661 CACCTCTCTTTCAGGAGAAATTAACAATGTTTATGGAGTGTGTAAGCCCATGGGATT 720
DB 806 CACCTCTCTTTCAGGAGAAATTAACAATGTTTATGGAGTGTGTAAGCCCATGGGATT 865
QY 721 GGGGAACAC-AATTCTTCGAAACCCAGATGCGGAGCAAAAGCCCTGGTCTTTTATTA 779
DB 866 GGGGRACACAATTTCTGCAGAAACCCAGATGCGGAGCAAAAG-CCTGGTCTTTTATTA 924
QY 780 AGTTACCAATCACAAGTGAATGGGAATATCTGTGATGCTCTCAGCTGCTCAGCCCGCA 839
925 AGTTACCAATCACAAGTGAATGGGAATATGTGATGCTCAGCTGCTCAGCCCGCA 984
940 CGTTGCTACCCAGAGAAAGCCCACTAGCCATCAACAGACTTCCGGGGTTTCACTC 899
985 CGTTGCTACCCAGAGAAAGCCCACTAGCCATCAACAGACTTCCGGGGTTTCACTC 1042

RESULT 2
BX463023 1008 bp mRNA linear EST 22-MAY-2003
LOCUS BX463023 Homo sapiens FETAL LIVER Homo sapiens cDNA clone
DEFINITION CS0DM004YK02 5-PRIME, mRNA sequence.
ACCESSION BX463023
VERSION BX463023.1 GI:31025478
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 1008)
AUTHORS Li, W.B., Gruber, C., Jesse, J. and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seq@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 5634.f for
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0DM004BF01QPI&cluster=5634.f. Contact :
Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ invitrogen Corporation 1600
Faraday Avenue Genoscope sequence ID : CS0DM004BF01QPI.

FEATURES
  source      Location/Qualifiers
1..1008
  /organism="Homo sapiens"
  /mol_type="mRNA"
  /db_xref="taxon:9606"
  /clone="CS0DM004YK02"
  /tissue="FETAL LIVER"
  /dev_stage="fetal"
  /clone_lib="Homo sapiens FETAL LIVER"
  /notes="Organ: liver; Vector: pCMVSPORT 6; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-strand cDNA was digested with Not I and
cloned into the Not I and EcoRV sites of the pCMVSPORT 6
vector. Library was not normalized."

ORIGIN
Query Match      45.6%; Score 767.4; DB 13; Length 1008;
Best Local Similarity 93.4%; Pred. No. 7.5e-208;
Matches 840; Conservative 31; Mismatches 19; Indels 9; Gaps 7;

QY 1 ATGTTTGCCAGAGTGTGATCTCCATGTTCTGCTTTAATGGCTCTGGTGGGAAAGACA 60
DB 110 ATGTTTGCCAGAGTGTGATCTCCATGTTCTGCTTTAATGGCTCTGGTGGGAAAGACA 169
QY 61 GCCTGTGGGTTCTCCCTGATGCTTTATTGGAAGCTCGGACCCAGACTGGACCCCTG-A 119
DB 170 GCCTGTGGGTTCTCCCTGATGCTTTATTGGAAGCTCGGACCCAGACTGGACCCCTGAA 229
QY 120 CCAGTATGATTACAGCTACGAGGATTAATATCAGGAAGAGAACACCAAGTAGCACATTAC 179
DB 230 CCAGTATGATTACAGCTACGAGGATTAATATCAGGAAGAGAACACCAAGTAGCACATTAC 289
QY 180 CCATGCT-GAGANCTCTGACTGCTACTACTAGTGGACCAAGCTGATCCATCCAGGCCA 238
DB 290 CCACCTAGAGAACTCTGACTGGTACTACTAGTGGACCAAGCTGATCCATCCAGGCCA 349
QY 239 ACCCTGTGAACACGGTGGGACTGCTGCTCCATGGGAGCACCTTCATGTCAGCTGCC 298
```

```

Db 350 ACCCTGTGAACAGCGTGGGACGTGCTGTCTCATGGGAAACCTTCACATCAGCTGCC 409
Qy 299 TGGCTCTTTCTCTGGGAATAAGT--TCAGAAAGTGCAAAATACGTGCAAGGCAACCC 356
Db 410 TGGCTCTTTCTCTGGGAATAAGTATCAFAHARATGCAAAATACGTGCAAGGCAACCC 469
Qy 357 ATGTGGCGGGGCCAATGT--CTCATACCCAGATCCTCCCTACTACTACCGTGTCTCTGT 414
Db 470 ATGTGGCGGGGCCAATGTACATCATATACCCAGATCCTCCCTACTACTACCGTGTCTCTGT 529
Qy 415 AAACACCTTACACAGTCCAG-CTGCTCCCAAGTGGTTCCTGTATGACGCGCAACCC 473
Db 530 AAACACCTTACACAGTCCAGTCTCCCAAGTGGTTCCTGTATGACGCGCAACCC 589
Qy 474 CTGCCAGATGGGGCTACCTGCTCCGGCATTAAGCCAGATCCAAAGTTCACCTGTGCTG 533
Db 590 CTGCCAGATGGGGCTACCTGCTCCGGCATTAAGCCAGATCCAAAGTTCACCTGTGCTG 649
Qy 534 TCCGACCACTTCAAGGGGAAATCTGTGAATAGTTCGTATGATCTGTATGTTGGGA 593
Db 650 TCMCGAMCAGMTCAAGGGGAAATCTGTGAATAGTTCGTATGATCTGTATGTTGGGA 709
Qy 594 TGGCTACTCTTACCGAGGAAATGAATAGGACAGTCAACACGATCGTGCCTTTACTG 653
Db 710 TGGCTACTCTTACCGAGGAAATGAATAGGACAGTCAACACGATCGTGCCTTTACTG 769
Qy 654 GAATCTCCACTCTCTTGAGGAGAAATTACAAATGTTTATGAGGATGCTGAAACCA 713
Db 770 GAATCTCCACTCTCTTGAGGAGAAATTACAAATGTTTATGAGGAGMTTAAACCA 829
Qy 714 TGGGATGGGGAAACAAATTTCTGCAGAAACCCAGATCGGACGAAAGCCCTGG-TGCT 772
Db 830 TGGGATGGGGAAACAAATTTCTGCAGAAACCCAGATCGGAGAAAGCCCTGTATGCT 889
Qy 773 TTATTAAGTACATGACAGGTGAAT-TGGGAATCTGTATGCTCAGCCTGCTCA 831
Db 890 TTATTAAGTACATGACAGGTGAAT-TGGGAATCTGTATGCTCAGCCTGCTCA 949
Qy 832 GCCCAGGACCTTGCTACCCAGAGGAAAGCCCACTGAGCCATCAACCAAGTTCCTGGG 890
Db 950 ACCASRACGKTGCTAMCCAGCGAGAGCCCACTGAGACATCAACCAARCTCCTGGG 1008

RESULT 3
BI761782
LOCUS 603046775f1 NIH_MGC_116 Homo sapiens cDNA clone IMAGE:5187066 5',
DEFINITION mRNA sequence.
ACCESSION BI761782
VERSION BI761782.1 GI:15753360
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NIH-MGC http://mgc.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM11467 row: e column: 19
High quality sequence stop: 839.
Location/Qualifiers
1..891

```

```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5187066"
/lab_host="DH10B"
/clone_lib="NIH_MGC_116"
/note="Organ: pooled colon, kidney, stomach; Vector:
pCMV-SPORT6; Site_1: NotI; Site_2: EcoRV (destroyed); RNA
source anonymous pool of 3 colons, age 26 yo male, 49 yo
female, 71 yo male colon, 46 yo male kidney, and pool of 2
stomachs, 62 yo male and 70 yo female. Library is
oligo-dT primed and directionally cloned (EcoRV site is
destroyed upon cloning). Average insert size 1.4 kb,
insert size range 1-3 kb. Library is normalized and
enriched for full-length clones and was constructed by C.
Gruber (Invitrogen). Research Genetics tracking code
023. Note: this is a NIH_MGC Library."

```

ORIGIN

```

Query Match 41.5%; Score 698.2; DB 12; Length 891;
Best Local Similarity 97.2%; Pred. No. 4.5e-188;
Matches 775; Conservative 0; Mismatches 13; Indels 9; Gaps 6;

Qy 1 ATGTTTGCAGATGTCGATCTCCATGTTCTGCTGTTTAAATGGCTCTGCTGGGAAAGACA 60
Db 81 ATGTTTGCAGATGTCGATCTCCATGTTCTGCTGTTTAAATGGCTCTGCTGGGAAAGACA 140
Qy 61 GCCTGTGGGTTCTCCCTGATGCTTTATTGGAAGCCCTGGAACCCAGACTGGACCCCTGAC 120
Db 141 GCCTGTGGGTTCTCCCTGATGCTTTATTGGAAGCCCTGGAACCCAGACTGGACCCCTGAC 200
Qy 121 CAGTATGATTACAGCTACGAGGATTAATCAGGAGAGAGACACACAGTAGCACACTTACC 180
Db 201 CAGTATGATTACAGCTACGAGGATTAATCAGGAGAGAGACACACAGTAGCACACTTACC 260
Qy 181 CATGCTGAAATCTCTGACTGTGTTACTACATGAGGACCAAGCTGATTCATGCGAGCCCAAC 240
Db 261 CAGCTGTAGAACTCTGACTGTGTTACTACATGAGGACCAAGCTGATTCATGCGAGCCCAAC 320
Qy 241 CCCTGTGAACAGGTGGGACCTGCTCTGTCATGAGGACCTTCAATGATGAGCTGCTG 300
Db 321 CCCTGTGAACAGGTGGGACCTGCTCTGTCATGAGGACCTTCAATGATGAGCTGCTG 380
Qy 301 GCTCCTTTCTCTGGGAATAAGTGTGAGAAAGTGCAAAATACGTGCAAGGACAAACCCATGT 360
Db 381 GCTCCTTTCTCTGGGAATAAGTGTGAGAAAGTGCAAAATACGTGCAAGGACAAACCCATGT 440
Qy 361 GGCGGGGCCCAATGCTCTCATTTACCCAGAGTCTCCCTACTACTACCGTGTGTGTAAACAC 420
Db 441 GGCGGGGCCCAATGCTCTCATTTACCCAGAGTCTCCCTACTACTACCGTGTGTGTAAACAC 500
Qy 421 CTTTACAGAGTCCAGCTGCTCCCAAGTGTTCCTGTATGAGGACCAACCCCTGCCAG 480
Db 501 CTTTACAGAGTCCAGCTGCTCCCAAGTGTTCCTGTATGAGGACCAACCCCTGCCAG 560
Qy 481 AATGGGGCTACCTGCTCCCGCATTAAGCGGAGATCCAAAGTTCACCTGTGCTGCTCCGAC 540
Db 561 AATGGGGCTACCTGCTCCCGCATTAAGCGGAGATCCAAAGTTCACCTGTGCTGCTCCGAC 620
Qy 541 CAGTTCAAGGGGAAATTTCTGTGAAAATAGTGTCTGATGATCTGATGTTGGGATGGCTAC 600
Db 621 CAGTTCAAGGGGAAATTTCTGTGAAAATAGTGTCTGATGATCTGATGTTGGGATGGCTAC 680
Qy 601 TCTTACCGAGGAAATTAATAGGACAGTCAACACGATCGTGCCTTTACTGGAACCT 659
Db 681 TCTTACCGAGGAAATTAATAGGACAGTCAACACGATCGTGCCTTTACTGGAACCT 740
Qy 660 CC-ACCTCTCTTTCAGGAGAAATTAACATGTTTATGGAGG---ATGCTGAAACCCATG 715
Db 741 CCAAGCTCTCTTTCAGGAGAAATTAACATGTTTATGGAGGATGCTTGAACCCCATG 800
Qy 716 GGATTCGGGAA--CACAAATTTCTGAG--AAACCCAGATGCGGAG--AAAACCCCTGCTGC 771

```

801 GGAATTGGGGAACACCAATTTCTGCAGAAAAACCCAGATGCCGACGAAAAAGCCCCCTGGTGC 860

772 TTTTATAAAGTTTACCAG 788

861 TTTTATAAAGTCAACCA 877

DB	241	GAGAATCTATTGAGGCGCTTTAAAGACACGGCGGGGAAGCACCCCATGTGCACGGCGTGTCUCCUCCA	300
QY	996	GTCTCGCTGCTCTGACCACTTCCTCCTCATGCCAGGGCCACTTCTGTGTGGGGGCTGTAT	1055
DB	301	GTCTCTGCTGCTCTGACCACTTCCTCCTCATGCCAGGGCCACTTCTGTGTGGGGGCTGTAT	360
QY	1056	CCACCCCCTGCTGGGTGCTCACTGTGTCGCCACTGCACCGACATAAAAAACGACATCTTAAA	1115
DB	361	CCACCCCCTGCTGGGTGCTCACTGTGTCGCCACTGCACCGACATAAAAAACGACATCTTAAA	420
QY	1116	GGTGTGCTTAGGGGACCAGGACCTTCAAGAAAGAAAGAAATTTTCATGACGACAGAGCTTTAGGGT	1175
DB	421	GGTGTGCTTAGGGGACCAGGACCTTCAAGAAAGAAAGAAATTTTCATGACGACAGAGCTTTAGGGT	480
QY	1176	GCAGAAGATATTCAAGTACAGCCACTACAATGAAGAGATGAGATTCCTCCACCAATGATAT	1235
DB	481	GCAGAAGATATTCAAGTTCAGTCACTACAATGAAGAGATGAGATTCCTCCACCAATGATAT	540
QY	1236	TGCATTGCTCAAGTTTAAAGCCAGTGGATGGTCACTGTCTCTAGAAATCCAAATACGTGAA	1295
DB	541	TGCATTGCTCAAGTTTAAAGCCAGTGGATGGTCACTGTCTCTAGAAATCCAAATACGTGAA	600
QY	1296	GACTGTGCTTGCCTGATGGGTGCTTTCCTCTCGGAGTGAAGTCCACAT-CTTGTGGCT	1354
DB	601	GACTGTGGGCTTGGCTGTGATGGTTCCTCTGTTAGAGAGTGCACATGTTGTGACT	660
QY	1355	GGGGTGTACAGAAACAGGAAAGGGTCCCGCAGCTCCTGGATGCCAAGTCAAAGCTGA	1414
DB	661	GTGGGTTCGAGAAACAGTTTAAGTTTCTTCCGCTACTGTTTCCCTAAGATAGTTCG	720
QY	1415	TTGCCAACACTTTGTGCAACTCCCGCCAACTCTATGACCACAT	1457
DB	721	TTGTCGACATTGGAGGGGGCGTGGGCTACTGTGTGTGCACAT	763
RESULT 5			
CB156834/c			
LOCUS			
DEFINITION K-EST0215807 L17N670205nl Homo sapiens cDNA clone			
L17N670205nl-4-F07 5', mRNA sequence.			
CB156834			
VERSION CB156834.1 GI:28141962			
KEYWORDS EST.			
SOURCE Homo sapiens (human)			
ORGANISM Homo sapiens			
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
AUTHORS 1 (bases 1 to 576)			
Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,			
Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and			
Kim,Y.S.			
TITLE 21C Frontier Korean EST Project 2001			
JOURNAL Unpublished (2002)			
COMMENT Contact: Kim YS			
Genome Research Center			
Korea Research Institute of Bioscience & Biotechnology			
52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea			
Tel.: +82-42-860-4470			
Fax: +82-42-860-4409			
Email: yongsung@mail.kribb.re.kr			
Plate: 4 row: F column: 07			
High quality sequence stop: 576.			
Location/Qualifiers			
1..576			
FEATURES			
source			
/organism="Homo sapiens"			
/mol_type="mRNA"			
/db_xref="taxon:9606"			
/clone="L17N670205nl-4-F07"			
/sex="F"			
/lab host="Top10F"			
/clone lib="L17N670205nl"			
/note="Organ: Liver; Vector: pt7T3-Pac; Site 1: EcoRI;			
sra 2: NORI: The library was contributed by the Soares			

```

801  GGATGGGGAACACCAATTTCTGCAGAAAACCCGATGCCGACGAAAACCCCTGTGC 860
772  TTTATTAAAGTTACCA 788
      |||||
861  TTTATTAAAGTCACCA 877

RESULT 4
BX431866      899 bp      mRNA      linear      EST 15-MAY-2003
BX431866 Homo sapiens FETAL LIVER Homo sapiens cDNA clone
CS0DM004YK02 5-PRIME, mRNA sequence.
BX431866
BX431866.1 GI:30789052
EST.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 899)
Li.W.B., Gruber,C., Jessee,J. and Polayes,D.
Full-length cDNA libraries and normalization
Unpublished (2001)
Contact: Genoscope
Genoscope - Centre National de Sequençage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 5634.f For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0BAG065ZH05_CS06184_i&cluster=5634.f.
Contact : Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Faraday Avenue Genoscope sequence ID : CS0BAG065ZH05_CS06184_1.
Location/Qualifiers
1. .889
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DM004YK02"
/tissue_type="FETAL LIVER"
/dev_stage="fetal"
/clone_lib="Homo sapiens FETAL LIVER"
/notes="Organ: liver; Vector: pCMVSPORT 6; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-strand cDNA was digested with Not I and
cloned into the Not I and EcoRV sites of the pCMVSPORT 6
vector. Library was not normalized."

ORIGIN
Query Match      38.3%; Score 645.4; DB 13; Length 889;
Best Local Similarity 91.2%; Pred. No. 5.9e-173;
Matches 696; Conservative
756  GGAGGATGCTGAACCCATGGGATGGGGACACAAATTCGAGAAACCCAGATCGGA 755
1  GGAGGATGCTGAACCCATGGGATGGGGACACAAATTCGAGAAACCCAGTTGCGGA 60
756  CGAAAGCCCTGGTGCCTTTTAAAGTTACCAATGACAAGTGGAATGGGAATACTGTGA 815
61  CGAAAGCCCTGGTGCCTTTTAAAGTTACCAATGACAAGTGGAATGGGAATACTGTGA 120
816  TGTCTAGCCTGCTACGCCAGACGTTGCTTACCAGAGAAAGCCCACTGAGCCATC 875
121  TGTCTAGCCTGCTCAGCCAGGACGTTGGCTACCCAGAGAAAGCCCCACTGAGCCATC 180
876  AACCAAGCTCCGGGGTTTGACTCCTGTGGAAGACTGAGATAGCAGAGAGAGATCAA 935
181  AACCAAGCTCCGGGGTTTGACTCCTGTGGAAGACTGAGATAGCAGAGAGAGATCAA 240
936  GAGAAATCTATGAGGCTTTAAGACACGCGCGGGAAGCACCCATGGCAGGGCTCCCTCCA 995

```


Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 736)

NIH-MGC <http://mgi.nci.nih.gov/>

National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999)

Contact: Robert Strausberg, Ph.D.
Email: cgabp@remail.nih.gov

Tissue Procurement: Dr. Michael Brownstein
cDNA Library Preparation: Michael Brownstein Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>

Plate: NDCW92 row: d column: 23
High quality sequence stop: 558.

FEATURES
source
1..736
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:30310462"
/lab_host="DH10B (Ti-phage-resistant)"
/clone_lib="NIH_MGC_177"
/notes="Organ: liver; Vector: pDNR-LIB; Site 1: SfiI (ggccattatggcc); Site 2: SfiI (ggcgccctggcc); cDNA made by oligo-dT priming and directionally cloned. 5' and 3' adaptors were used in cloning as follows:
5'-AAGCAGGTGATACGAGGAGGCGGATGACGGCCGG-3' and
5'-ATTCTAGGCGGAGGCGGCGGATG-3' (30)NN-3'. Full-length enriched library was constructed using the Clontech Creator SMART kit and size-selected to contain the 0.5 kb size fraction. Library created in the laboratory of M. Brownstein (NIMH, NIH). Note: this is a NIH_MGC Library."

ORIGIN
Query Match 27.0%; Score 454.2; DB 14; Length 736;
Best Local Similarity 80.8%; Pred. No. 2.9e-118;
Matches 542; Conservative 0; Mismatches 128; Indels 1; Gaps 1;

QY 715 GGGATTCGGGACACAACTTCTGAGAAACCCAGATGCGGACGAAAGCCCTGGTCTTT 774
DB 3 GGGATTCGGGACACAACTTCTGAGAAACCCAGATGCGGACGAAAGCCCTGGTCTTT 52
QY 775 ATTAAGTTACCAATGACAAGGTGAATGGGAATCTGTGATCTCAGCCTGCTCAGCC 834
DB 63 GTCAAGGTGAACAGTGAAGAGTGAATGGGAATCTGTGATCTCAGCCTGCTCAGTG 122
QY 835 CAGGACGTTGCTTCCACGAGGAGGAGCCCACTGAGCCATCACCAAGCTTCGGGGTTT 894
DB 123 CTTGACACCCCTAACCCAGTGGAAAGCCTTCTGGAGCCTGTGATGGAGCTGCCAGGTTT 182
QY 895 GACTCTGTGGAAGACTGAGATAGCAGAGGAGGAGATCAAGAGAAATCTATGAGGCTTT 954
DB 183 GAGTCTTGGGAGAGCGAGGTAGTGAACACGAGTCAAGCGTATCTACGGGGCTTT 242
QY 955 AAGAGCAGGGGCGGAGCAGCCATGCGAGGCTCCCTCCAGTCCCTGCTGCTCTGACC 1014
DB 243 AAGAGCAGGAGGAGCAGCCCGTGGCAGGTGTCCTGAGACCTTCTGAGCTTGAAC 302
QY 1015 ATCTCCATGCCCCAGGCGCACTTCTGTGTGGGGGCTGTATCCACCCCTCTGGGTGCTC 1074
DB 303 ACCTCCATGCCCCAGGCGCACTTCTGTGTGGGGGCGCTGTATCCACCCCTCTGGGTGCTC 362
QY 1075 ACTGCTGCCACTCTACCGACATATAAACACAGACATCTAAGGTGGTGTAGGGGACCAG 1134
DB 363 ACTGACGCCACTCTATCCGACATATAAACACAGACATCTAAGGTGGTGTAGGGGATCAG 422
QY 1135 GACCTGAAGAAAGAAATTTTCATGAGCAGAGCTTTTAGGGTGCAGAGATATTCAAGATC 1194
DB 423 GACCTGAAGAAAGACAGATCCCATGACAGACCTTCAGGGTGGAAATACTGAAGTAC 482

QY 1195 AGGCATACATGAAGAGATGAGATTCGCCACAAATGATATTCATTCCTCAAGTTAAAG 1254
DB 483 AGTCAGTATATGAAGAGATGAGATTCGCCACAAATGATATTCCTTGTCTCAAGTTAAAG 542

QY 1255 CCAGTGCATGCTCACTGTGCTCTAGAAATCCAAATAGCTGAAGACTGTGCTTGCCTGAT 1314
DB 543 CCAGTGGTGGTCACTGTGCTCTGGATCCAGATATGTGAGACTGTATGTTGCCAGC 602

QY 1315 GGGTCTTTTCCCTCTGGAGTGAAGTGCACATCTTGGCT-GGGTGTACAGAAACAG 1373
DB 603 GACCCCTTTCCCTCTGGAGTGAAGTGCACATCTTGGCTGGGGGTGTTTCAGAAACAG 662

QY 1374 AAAAGGGTCCC 1384
DB 663 GGAAGGGGTCC 673

RESULT 8
BE032018
LOCUS BE032018 592 bp mRNA linear EST 09-JUL-2000
DEFINITION 130809 MARC 1P1G Sus scrofa cDNA 5', mRNA sequence.
ACCESSION BE032018
VERSION BE032018.1 GI:8327027
KEYWORDS EST.
SOURCE Sus scrofa (pig)
ORGANISM Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
1 (bases 1 to 592)
Fahrenkrug, S.C., Smith, T.P.L., Freking, B.A., Cho, J., White, J.,
Vallet, J., Wise, T., Rohrer, G.A., Perlee, G., Sultana, R.,
Quackenbush, J. and Keele, J.W.
Porcine gene discovery by normalized cDNA-library sequencing and
EST cluster assembly
Mamm. Genome 13 (8), 475-478 (2002)
22213789
12226715
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and alt trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 18
and -minmatch 12 options.
PCR Primers
FORWARD: AGGAACAGCTATGACCAT
BACKWARD: GTTTCACGTCACGACG
Plate: 69 row: K column: 12
Seq primer: ATTTAGGTGACACTATAG.
Location/Qualifiers
1..592
/organism="Sus scrofa"
/mol_type="mRNA"
/db_xref="taxon:9823"
/tissue_type="pooled"
/lab_host="DH10B"
/clone_lib="MARC 1P1G"
/note="Vector: pCMV SPORT6; Site 1: NotI; Site 2: SalI;
Library made from pooled tissue from day 11, 13, 15, 20,
and 30 embryos."

ORIGIN
Query Match 26.8%; Score 451.6; DB 10; Length 592;
Best Local Similarity 86.0%; Pred. No. 1.4e-117;
Matches 518; Conservative 0; Mismatches 69; Indels 15; Gaps 1;

QY 522 CACCTGTGCTGTCGCCAGCTTCAAGGGGAAATCTGTGAAATAGGTTCTGATGACTG 581
DB 6 CGCGTGGGCTGCTCTGACCAAGTTTAAAGGGGAGATCTGTGAAATAGGTTCTGATGACTG 65
QY 592 CTATCTGGCGTGGCTACTCTTCCGAGGGGAAATGAATAGGACAGTCAACACGATGC 641

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 736)

NIH-MGC <http://mgi.nci.nih.gov/>

National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999)

Contact: Robert Strausberg, Ph.D.
Email: cgabp@remail.nih.gov

Tissue Procurement: Dr. Michael Brownstein
cDNA Library Preparation: Michael Brownstein Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>

Plate: NDCW92 row: d column: 23
High quality sequence stop: 558.

FEATURES
source
1..736
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:30310462"
/lab_host="DH10B (Ti-phage-resistant)"
/clone_lib="NIH_MGC_177"
/notes="Organ: liver; Vector: pDNR-LIB; Site 1: SfiI (ggccattatggcc); Site 2: SfiI (ggcgccctggcc); cDNA made by oligo-dT priming and directionally cloned. 5' and 3' adaptors were used in cloning as follows:
5'-AAGCAGGTGATACGAGGAGGCGGATGACGGCCGG-3' and
5'-ATTCTAGGCGGAGGCGGCGGATG-3' (30)NN-3'. Full-length enriched library was constructed using the Clontech Creator SMART kit and size-selected to contain the 0.5 kb size fraction. Library created in the laboratory of M. Brownstein (NIMH, NIH). Note: this is a NIH_MGC Library."

ORIGIN
Query Match 27.0%; Score 454.2; DB 14; Length 736;
Best Local Similarity 80.8%; Pred. No. 2.9e-118;
Matches 542; Conservative 0; Mismatches 128; Indels 1; Gaps 1;

QY 715 GGGATTCGGGACACAACTTCTGAGAAACCCAGATGCGGACGAAAGCCCTGGTCTTT 774
DB 3 GGGATTCGGGACACAACTTCTGAGAAACCCAGATGCGGACGAAAGCCCTGGTCTTT 52
QY 775 ATTAAGTTACCAATGACAAGGTGAATGGGAATCTGTGATCTCAGCCTGCTCAGCC 834
DB 63 GTCAAGGTGAACAGTGAAGAGTGAATGGGAATCTGTGATCTCAGCCTGCTCAGTG 122
QY 835 CAGGACGTTGCTTCCACGAGGAGGAGCCCACTGAGCCATCACCAAGCTTCGGGGTTT 894
DB 123 CTTGACACCCCTAACCCAGTGGAAAGCCTTCTGGAGCCTGTGATGGAGCTGCCAGGTTT 182
QY 895 GACTCTGTGGAAGACTGAGATAGCAGAGGAGGAGATCAAGAGAAATCTATGAGGCTTT 954
DB 183 GAGTCTTGGGAGAGCGAGGTAGTGAACACGAGTCAAGCGTATCTACGGGGCTTT 242
QY 955 AAGAGCAGGGGCGGAGCAGCCATGCGAGGCTCCCTCCAGTCCCTGCTGCTCTGACC 1014
DB 243 AAGAGCAGGAGGAGCAGCCCGTGGCAGGTGTCCTGAGACCTTCTGAGCTTGAAC 302
QY 1015 ATCTCCATGCCCCAGGCGCACTTCTGTGTGGGGGCTGTATCCACCCCTCTGGGTGCTC 1074
DB 303 ACCTCCATGCCCCAGGCGCACTTCTGTGTGGGGGCGCTGTATCCACCCCTCTGGGTGCTC 362
QY 1075 ACTGCTGCCACTCTACCGACATATAAACACAGACATCTAAGGTGGTGTAGGGGACCAG 1134
DB 363 ACTGACGCCACTCTATCCGACATATAAACACAGACATCTAAGGTGGTGTAGGGGATCAG 422
QY 1135 GACCTGAAGAAAGAAATTTTCATGAGCAGAGCTTTTAGGGTGCAGAGATATTCAAGATC 1194
DB 423 GACCTGAAGAAAGACAGATCCCATGACAGACCTTCAGGGTGGAAATACTGAAGTAC 482

QY 1195 AGGCATACATGAAGAGATGAGATTCGCCACAAATGATATTCATTCCTCAAGTTAAAG 1254
DB 483 AGTCAGTATATGAAGAGATGAGATTCGCCACAAATGATATTCCTTGTCTCAAGTTAAAG 542

QY 1255 CCAGTGCATGCTCACTGTGCTCTAGAAATCCAAATAGCTGAAGACTGTGCTTGCCTGAT 1314
DB 543 CCAGTGGTGGTCACTGTGCTCTGGATCCAGATATGTGAGACTGTATGTTGCCAGC 602

QY 1315 GGGTCTTTTCCCTCTGGAGTGAAGTGCACATCTTGGCT-GGGTGTACAGAAACAG 1373
DB 603 GACCCCTTTCCCTCTGGAGTGAAGTGCACATCTTGGCTGGGGGTGTTTCAGAAACAG 662

QY 1374 AAAAGGGTCCC 1384
DB 663 GGAAGGGGTCC 673

RESULT 8
BE032018
LOCUS BE032018 592 bp mRNA linear EST 09-JUL-2000
DEFINITION 130809 MARC 1P1G Sus scrofa cDNA 5', mRNA sequence.
ACCESSION BE032018
VERSION BE032018.1 GI:8327027
KEYWORDS EST.
SOURCE Sus scrofa (pig)
ORGANISM Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
1 (bases 1 to 592)
Fahrenkrug, S.C., Smith, T.P.L., Freking, B.A., Cho, J., White, J.,
Vallet, J., Wise, T., Rohrer, G.A., Perlee, G., Sultana, R.,
Quackenbush, J. and Keele, J.W.
Porcine gene discovery by normalized cDNA-library sequencing and
EST cluster assembly
Mamm. Genome 13 (8), 475-478 (2002)
22213789
12226715
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and alt trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 18
and -minmatch 12 options.
PCR Primers
FORWARD: AGGAACAGCTATGACCAT
BACKWARD: GTTTCACGTCACGACG
Plate: 69 row: K column: 12
Seq primer: ATTTAGGTGACACTATAG.
Location/Qualifiers
1..592
/organism="Sus scrofa"
/mol_type="mRNA"
/db_xref="taxon:9823"
/tissue_type="pooled"
/lab_host="DH10B"
/clone_lib="MARC 1P1G"
/note="Vector: pCMV SPORT6; Site 1: NotI; Site 2: SalI;
Library made from pooled tissue from day 11, 13, 15, 20,
and 30 embryos."

ORIGIN
Query Match 26.8%; Score 451.6; DB 10; Length 592;
Best Local Similarity 86.0%; Pred. No. 1.4e-117;
Matches 518; Conservative 0; Mismatches 69; Indels 15; Gaps 1;

QY 522 CACCTGTGCTGTCGCCAGCTTCAAGGGGAAATCTGTGAAATAGGTTCTGATGACTG 581
DB 6 CGCGTGGGCTGCTCTGACCAAGTTTAAAGGGGAGATCTGTGAAATAGGTTCTGATGACTG 65
QY 592 CTATCTGGCGTGGCTACTCTTCCGAGGGGAAATGAATAGGACAGTCAACACGATGC 641

```

66 CTATGTTGGTGACGGCTACTCTTACCAGGGGGAAGTGAATAAACTGTCAACAGCACAC 125
642 GTGCTTACTGGAACCTCCACTCTTCCAGAGAAATTAACAATGTTTATGGAGGA 701
126 GTGCTTACTGGAACCTCCACTCTTCCAGAGAAATTAACAATGTTTATGGAGGA 185
702 TGCTGAACCCATGGATGGGGAACAAATTTCTGCAGAAACCCAGATGGGACGAAAA 761
186 TGCTGAGCCCATGGATGGGAGGACAACTTCTGCAGAAACCCAGATGGAGACAGAA 245
762 GCTCTGGTGTCTTATTAAGTTACCAATGACAAGTGAAATGGGAATACTGTGATCTC 821
246 GCTCTGGTGTCTTATTAAGTTACCAATGACAAGTGAAATGGGAGTACTGTGACGTCC 305
822 AGCTGCTCAGCCAGGAGCTTGCTACCCAGAGAAAGCCCACTGAGCCATCAACAA 881
306 TGCTGCTCAGC-----CCAGAGGAGAGCCCACTAGAACCTGTGACCGA 350
882 GCTTCCGGGGTTGACTCTCTGTGGAAGACTGAGATAGCAGAGAGAAATCAAGAGAT 941
351 GCTTCCGGGGTTGACTCTCTGTGGAAGACTGAGATAGCAGAGAGAAATCAAGAGAT 410
942 CTATGAGGCTTTAGACAGCGCGGCAAGCAGCCATGCGAGGCGTCCCTCCAGTCTC 1001
411 CTACGAGGCTTTAAGACAGCGCGGCAAGCAGCCATGCGAGGCGTCCCTCCAGTCTC 470
1002 GCTGCTCTGACCATCTCCATGCCAGGCGGCAAGCAGCCATGCGAGGCGTCCCTCCAGTCTC 1061
471 CTTCGATCTGACCGTCTCCATGCCAGGCGGCAAGCAGCCATGCGAGGCGTCCCTCCAGTCTC 530
1062 CTGCTGGTGTCTCACTGTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCT 1121
531 CTGCTGGTGTCTCACTGTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCT 590
1122 GC 1123
591 GC 592

```

```

RESULT 9
BG972579
LOCUS
DEFINITION
5', mRNA sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

```

BG972579 802 bp mRNA linear EST 12-JUN-2001
 602841165F1 NCI CGAP_Kid14 Mus musculus cDNA clone IMAGE:4975384
 5', mRNA sequence.
 BG972579.1 GI:14360216
 EST.
 Mus musculus (house mouse)
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 802)
 NIH-MGC <http://mgs.nci.nih.gov/>;
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-x@mail.nih.gov
 Tissue Procurement: Jeffrey E. Green, M.D.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: ILAM10967 row: a column: 17
 High quality sequence stop: 785.
 Location/Qualifiers
 1..802
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="FVB/N"
 /db_xref="taxon:10090"

```

/clon="IMAGE:4975384"
/lab host="DH10B (T1 phage-resistant)"
/clon lib="NCI CGAP Kid14"
/Note="Organ: Kidney; Vector: pCMV-Sport6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.75 kb. Constructed by Life
Technologies. Note: this is a NCI CGAP Library."

```

Query Match	Score	447.2	DB 12	Length	802
Best Local Similarity	77.7%	Pred. No. 3.1e-116			
Mismatches	580	Conservative	0	Mismatches	158
				Indels	8
				Gaps	3
QY	1	AGTTTGCAGAGATGTCATCTCCATGTTCTGCTGTTAATGGCTCTGTTGGGAAGACA	60		
DB	50	ATATTTGTCCAGATGTTGTTGTTCCGTGCTTAAATCGCTGTTGGGGAAGTCA	109		
QY	61	GCTGTGGGTCTCCCTGATGTTTATTGGAAGCCTGACCCAGACCTGGACCCCTGAC	120		
DB	110	GTCAATGGGCTCTCACTGATGTTCTTCAATGCGCCCGCCAGACCCAGATGGACCCCGAT	169		
QY	121	CAGTATGATTACAGCTACGAGGATTAATATAGGAAGAAACACAGTAGACACTTACC	180		
DB	170	GACTATTACTACAGCTATGAGCAGTCCAGCCAGAGAGCCCAAGTGTACGAGAGCC	229		
QY	181	CATGCTGAGAACTCTGACTGTTACTACTGAGGACCAAGCTGATCCATGCCAGCCCAAC	240		
DB	230	ACCCCTGAGAAACCCCGACTGTTACT-----ATGAGACGATGATCCATGCCAGTCCAAC	283		
QY	241	CCCTGTGAACACGGTGGGACTGCTGCTGCTCATGGAGACACCTTACATGACAGTGCCTG	300		
DB	284	CCCTGTGAACACGGGCGGACTGTATCATCAGAGGGGATACCTTCAAGTGTGCAAGTCCCA	343		
QY	301	GCTCTCTTCTCTGGGAATTAAGTGTGAGAAAGTCAAAATACGTGCAAGGACCAACCCATGT	360		
DB	344	GCCCCCTTCTCGGGGAGCCGTTGCCAGACTGCAAAACAGTGCAGAGACCAACCCATGT	403		
QY	361	GGCGGGGCGCAATGTCTCATTTACCCAGAGTCTTCCCTACTACGCTGTGTCTTAAACAC	420		
DB	404	GTCCATGTTGATTTGCTTCTCATTTACCCAGAGACCCCTTACTACGCTGTGTCTTAAATAC	463		
QY	421	CCTTACAGAGTCCAGCTGCTCCAGTGTTCCTGTATGAGGCGCAACCCCTGCCAG	480		
DB	464	CCTTACAGGAGCCAGACTGCTCCAAAGTGTCTCCGCGATGAGGCGCAACCCCTGCCAG	523		
QY	481	AATGGGCTACCTGCTCCCGGCTAAGCGGAGATCCAAGTTCACCTGTGCTTCCCGAC	540		
DB	524	AATGGCGAGTCTGTTCCCGACACAGACGAGATCCAGGTTTACCTGTGCTTCCAGAC	583		
QY	541	CAGTTCAGGGGAAATTTCTGTA-ATAGTTTCTGATGCTGCTATGTCGCGATGGCTA	599		
DB	584	CAGTATAAGGGGAAATTTCTGTAACATAGTTCGCGAGCTGTTATGTCTGCGTATGGCTA	643		
QY	600	CTCTTACCGAGGAAATGAATAGGACAGTCAACCCAGCATGCGTCCCTTTACTGGAACTC	659		
DB	644	CTCTTACCGAGGCAAGTGAAGACAGTCAACCCAGCATGCGTCCCTTTACTGGAACTC	702		
QY	660	CCACTCTCTTTCAGGAGAAATTAACATGTTTATGGAGGATGCTGAACCCATGGAT	719		
DB	703	CCAACTCTCTTTCAGGAGAAATTAACATGTTTATGGAGGATGCTGAACCCATGGAT	762		
QY	720	TGGGGAACACAAATTTCTGCAGAAACC	745		
DB	763	GCGAGACCAACTTCTGCAGAAACC	788		

AA217892 674 bp mRNA linear EST 06-FEB-1997
 IMAGE:658839 5. similar to TR:G1345398 G1345398 HGF ACTIVATOR LIKE
 PROTEIN. i, mRNA sequence.
 AA217892

```

RESULT 10
AA217892
LOCUS
DEFINITION
ACCESSION

```

AA217892.1 GI:1826875
Mus musculus (house mouse)
EST.
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Possible reversed clone: similarity on wrong strand
Seq primer: -28ml3 rev2 BT from Amersham
High quality sequence stop: 485.
Location/Qualifiers
1. 674
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:658839"
/sex="unknown"
/tissue_type="fetus"
/dev_stage="12.5dpc total fetus"
/lab_host="DH103"
/clone_lib="Soares mouse 3NME12 5"
/note="Organ: whole fetus; Vector: pTV73D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', TGTTACCACTGAGTGGAGCGCGCGCTTATTTTATTTTATTTT 3'], on total mouse RNA [provided by Minoru Ko, Wayne State Univ.]; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pTV73 vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo."

Db 241 TCTACGGGGCTTTAAGACGACAGCAGGCAAGCAGCCCGTGGCAGGTGTCTCCCTGCAGACCT 300
Qy 1001 CGCTGCTCTGACCATCTCATGCCCGCAGGGCCACTTCTGTGTGGGGCGCTGATCCACC 1060
Db 301 CACTGCGGTTGACCACTCCATGCCCAAGGCCACTTCTGTGGGGCGCGCTGATCCACC 360
Qy 1061 CCTGCTGGGTGCTCACTGTGCCCACTGCACCGACATAAAACCAACACATCTAAAGGTGG 1120
Db 361 CCTGCTGGGTGCTCACTGTGCCCACTGTACCGACATAACCAACAGCATCTAAAGTTG 420
Qy 1121 TGCTAGGGGACCAAGGACCTGGAAGAAAGAAATTTATGAGCAGAGCTTTAGGGTGCAGA 1180
Db 421 TACTAGGGGATCAGGACCTGGAAGACAGAGATCCCATGACAGACCTTCAGGGTGGAAA 480
Qy 1181 AGATATTCAGTACAGCCACTCAATGAAGAGATGAGATCCCAATGATATTCAT 1240
Db 481 AATATCTGAAGTACAGTCACTATATGAAGAGATGAGATTTCCCAATGATATTCAT 540
Qy 1241 TGCTCAAGTTAAAGCCAGTGGTGCATCTGTCTCTAGAAATCAATACGTGAAGACTG 1300
Db 541 TGCTCAAGTTAAAGCCAGTGGTGCATCTGTCTCTGTGAATCCAGATATGTGAAGACTG 600
Qy 1301 TGCTCTTGCCTGATGGTGCCTTCCCTCTGGAGTGTAGTCCACATCTCTGGCTGGGGTG 1360
Db 601 TATGTTTGCACAGGACCCCTTTGCCCTCTGGACTGAGTGCACATCTCTGGACTGGGTG 660
Qy 1361 TTACAGAAACAGG 1373
Db 661 GTACAGAAACAGG 673

RESULT 11
BI148082
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

BI148082 751 bp mRNA linear EST 05-JUL-2001
602912432F1 NCI_CGAP_L19 Mus musculus cDNA clone IMAGE:5053663 5',
mRNA sequence.
BI148082.1 GI:14608083
EST.
Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 751)
NIH-MGC http://mgi.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Inyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM1145 row: 0 column: 08
High quality sequence stop: 747.
Location/Qualifiers
1. 751
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:5053663"
/lab_host="DH10B (T1 phage-resistant)"
/clone_lib="NCI CGAP L19"
/note="Organ: liver; Vector: pCMV-SPORT6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.9 kb. Constructed by Life
Technologies. Note: this is a NCI_CGAP Library."

FEATURES
source
ORIGIN

AA217892.1 GI:1826875
Mus musculus (house mouse)
EST.
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Possible reversed clone: similarity on wrong strand
Seq primer: -28ml3 rev2 BT from Amersham
High quality sequence stop: 485.
Location/Qualifiers
1. 674
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:658839"
/sex="unknown"
/tissue_type="fetus"
/dev_stage="12.5dpc total fetus"
/lab_host="DH103"
/clone_lib="Soares mouse 3NME12 5"
/note="Organ: whole fetus; Vector: pTV73D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', TGTTACCACTGAGTGGAGCGCGCGCTTATTTTATTTTATTTT 3'], on total mouse RNA [provided by Minoru Ko, Wayne State Univ.]; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pTV73 vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo."

Query Match 25.9%; Score 436.2; DB 9; Length 674;
Best Local Similarity 78.0%; Pred. No. 4e-113;
Matches 525; Conservative 0; Mismatches 148; Indels 0; Gaps 0;

Qy 701 ATGCTGAACCCATGGGATGGGACACACATTTCTCAGAAACCCAGATGGCGAGCAA 760
Db 1 ATTGGGAAGCGCGGATCGCAGACACACTTCTCAGAAACCCAGATGGGAGCCACA 60
Qy 761 AGCCCTGTGCTTTTATTAAGTTACCAATGACAGGTGAATGGGAATCTGTATGCT 820
Db 61 AACCTGTGTTTCTGCTCAAGGTGACAGTGAAGAGTGAATGGGAATCTGTATGCTCA 120
Qy 821 CAGCTGTCTAGCCAGAGAGCTTGCCTACCCAGAGGAAAGCCCACTGAGCATCAACCA 880
Db 121 CAGTGTGCTAGTCTGACACCCCTAACCCAGTGGAAAGCTTCTGGAGCTGTGTGG 180
Qy 881 AGCTTCGCGGGTTTGACTCTCTGTGGAAGACTGAGATAGCAGAGAGGAAGATCAAGAA 940
Db 181 AGCTCCAGGGTTTCAGTCTCTGCGGAAGACGGAGTGTAGTGAACCGCAGTCAAGGTA 240
Qy 941 TCTATGGAGGCTTTAAGAGCAGCGCGGCGAGACCCCATGGCAGGCTCCCTCCAGTCCT 1000

	Query Match	25.7%	Score 432;	DB 12;	Length 570;
	Best Local Similarity	85.0%;	Pred. No. 5.7e-112;	Indels	0; Gaps
	Matches 483; Conservative	0;	Mismatches 89;		
QY	1108	CATCTAAAGGTGGTGTCTAGGGGACCGAGGACCTGTGAAGAAAGAGAAATTTCTATGACGAGC	1167		
DB	3	CATCTAAAGGTTGTACTAGGGATCAGGACCTGTGAAGAGACAGAAATCCCATGAACAGACC	62		
QY	1168	TTTAGGCTGCAGAGATATTTCAAGTACAGCCACTACATGAAGAGATGAGATTTCCCCAC	1227		
DB	63	TTCAGGTGGAAATAATCTGAAGTACAGTCAGTAGTATATGAAGAGATGAGATTTCCCCAC	122		
QY	1228	AATGATATTGATTTGCTCAAGTTAAAGCCAGTGGATGGTCACTGTGCTCTAGATCCAAA	1287		
DB	123	AATGACATTGCTTTGGTCAAGTTAAAGCCAGTGGGTGGTCACTGTGCTCTGGAATCCAGA	182		
QY	1288	TACGTGAAGACTGTGTGCTGTGCCTGATGGGTCTTTCCCTCTGGGAGGTGAGTGGCCATC	1347		

Db 183 TATGTGAAGACTGTATGTTTGGCCAGAGACCCCTTCCCTCTGGAAGTGAAGTGCACATC 242
 QY 1348 TGTGGCTGGGGTGTATACAGAAACAGAAAGGGTCCGCCAGCTCTCGGATGCCAAAGTC 1407
 Db 243 TGTGGCTGGGGTGTATACAGAAACAGGGAAGGGTCCGCCAGCTCTCGGATGCCAAAGTC 302
 QY 1408 AAGCTGATTTGCCAACACATTTGTGCAACTCCGCCAACTCTATGACCAATGATGATGAC 1467
 Db 303 AAGCTAATCGCTAACCCCTTTGTGCACTCCGCCAACTCTATGACCAATGATGATGAC 362
 QY 1468 AGTATGATCTGTGCAAGGAATCTTACAGAACTCTGGCAAGACACCTGCCAGGGTGAATCT 1527
 Db 363 AGTATGATTTGTGCGGGGAACCTTCAGAAAGCCGAGTACAGACACCTGCCAGGGTGAATCT 422
 QY 1528 GGAGGGCCCTGACCTGTGAGAAGGACCGCACTACTACTAGCTCTATGGATAGTGAAGTGG 1587
 Db 423 GGGGGCCCTCTAACCTGTGAGAAGGATGGAATCTACTAGCTCTACGGATGTAAGTGG 482
 QY 1588 GGCCTGGAGTGTGAGAAGGACCGGGTCTTACACCAAGTTACAAATTCCTGAATGG 1647
 Db 483 GGCCAGGAATGTGGGAAGAGCAGGAGTCTACACTCAAGTCACCAAGTCTCTGAATGG 542
 QY 1648 ATCAAAGCCACCATCAAAAGTCAAAAGTG 1675
 Db 543 ATAAAGACCACCATGCACAGGAGGCTG 570

RESULT 13

BF788188 929 bp mRNA linear EST 12-JAN-2001
 LOCUS 602113411F1 NCI_CGAP_Kid14 Mus musculus cDNA clone IMAGE:4241642
 DEFINITION 5', mRNA sequence.

ACCESSION BF788188

VERSION BF788188

KEYWORDS EST.

SOURCE BF788188.1 GI:12093224

ORGANISM Mus musculus (house mouse)

Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

1 (bases 1 to 929) Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

TITLE NIH-MGC <http://mgc.nci.nih.gov/>

JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)

COMMENT Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgaps-remail.nih.gov

Tissue Procurement: Jeffrey E. Green, M.D.

cDNA Library Preparation: Life Technologies, Inc.

DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: LNA9858 row: e column: 03

High quality sequence stop: 686.

Location/Qualifiers

1..929

/organism="Mus musculus"

/mol_type="mRNA"

/strain="FVB/N"

/db_xref="taxon:10090"

/clone="IMAGE:4241642"

/lab_host="DH10B (T1 phage-resistant)"

/clone_lib="NCI_CGAP_Kid14"

/note="Organ: kidney; Vector: pCMV-SPORT6; Site 1: NotI;

Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt.

Average insert size 1.75 kb. Constructed by Life

Technologies. Note: this is a NCI_CGAP Library. |"

ORIGIN

Query Match 24.7%; Score 416.4; DB 10; Length 929;
 Best Local Similarity 75.9%; Pred. No. 2.3e-107;
 Matches 529; Conservative 0; Mismatches 162; Indels 6; Gaps 1;

QY 1 AGTTTGCAGGATGTTCTGATCTCCATGTTCTGCTGTTAATGCTCTGTTGGGAAAGACA 60
 Db 44 ATATTGTGTCAGGATGTAGGTGTTCCGTTGCTCTGCTGCTAATTCGCCCTTGGTGGGAAAGTCA 103
 QY 61 GCCTGTGGGTCTCCCTGATGTTTATTGGAAAGCCTGACCCAGACAGCTGGACCCCTGAC 120
 Db 104 GTCATGGGGCTCTCACTGATGTTCTTATTCGCCCCCAGACCCAGATGGACCCCGAT 163
 QY 121 CAGTATGTTACACTAGGAGGATTAATCAGAAAGAGAACACCACTAGTACACTTACC 180
 Db 164 GACTATTACTACACTATGAGCAGTCCAGCCCAAGAGACCCCACTGTCACGAGAC 223
 QY 181 CATGCTGAGAACTCTGATCTGTACTACACTGAGGACCAAGCTGATCCATGCCAGCCCAAC 240
 Db 224 ACCCTGAGAACCCCGACTGTACT-----ATGAAGACGATGATCCATGCCAGTCCAAC 277
 QY 241 CCCTGTGAACCGTGGGACTGCTGTCATCGGAGCACCTTTCATGCACTGACCTGCTG 300
 Db 278 CCCTGTGAACCGCGGGGACTGTATCATCAGAGGGGATACCTTTCAGTTGCACTGCCCA 337
 QY 301 GCTCCTTTCTCTGGGAATAAGTGTTCAGAAAGTGCAGAAATACGTGCAAGGACCAACCCATGT 360
 Db 338 GCCCCTTTCTCGGGAGCCGTCAGACTGCACAAACCAAGTGAAGGACCAACCCATGT 397
 QY 361 GCGCGGGCCAAATGTTCTCATTAACCCAGAGTCTCCCTACTACCGCTGTGCTGTAACAC 420
 Db 398 GTCCATGTGATTGCTTCATTAACCCAGAGACCCCTACTACCGCTGTGCTGCAATAC 457
 QY 421 CTTTACAGAGTCCCACTGCTCCCAAGTGGTTCTGTATGTCAGAGGCAAAACCCCTGCCAG 480
 Db 459 CTTTACAGGACAGACTGCTCCAAAGTCTTCCGCGATGCAGGCAAAACCCCTGCCAG 517
 QY 481 AATGGGCTACTCTCCCGCATAGCGGAGATCCAGTTACCTGCTGCTGCTGCTGCTGCTG 540
 Db 518 AATGGGAGTCTGTTCCCGACACAGCGAGATCCAGTTTACCTGCTGCTGCTGCTGCTG 577
 QY 541 CAGTTTCAAGGGGAAATCTGTGAATAGTTTCTGATGACTGCTGATGTTGGCGATGCTAC 600
 Db 578 CAGTATAAGGGGAAATCTGTGAATAGTCCGAGACTGTTATGTCGCTGATGCTGCTAC 637
 QY 601 TCTTACCGAGGAAATGATAGACAGTCAACAGCATGCGTGCCTTTACTTGGAACTCC 660
 Db 638 TCTTACCGAGGAAAGTGAAGACAGTCCACAGAACCACTGCTTACTCGGACTCC 697
 QY 661 CACTCTCTTTCGAGGAGAAATTACACATGTTTATGG 697
 Db 698 CACTCTCTTTCGAGGAGACTTATCATGTTGCTCTCG 734

CF169114 665 bp mRNA linear EST 25-JUL-2003
 B0809C08-5 NIA Mouse Newborn Kidney cDNA Library (Long 1) Mus
 musculus cDNA clone NIA:B0809C08 IMAGE:30468319 5', mRNA sequence.

ACCESSION CF169114

VERSION CF169114.1

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

1 (bases 1 to 665) Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

TITLE Piao, Y., Ko, N.T., Lim, M.K. and Ko, M.S.H.

AUTHORS Construction of long-transcript enriched cDNA libraries from

submicrogram amounts of total RNAs by a universal PCR amplification

method

JOURNAL Genome Res. 11 (9), 1553-1558 (2001)

MEDLINE 21429098

PUBMED 11544199

COMMENT Contact: Dawood B. Dudekula

Laboratory of Genetics

National Institute on Aging/National Institutes of Health

333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA

Email: cdna@lgsun.grc.nia.nih.gov
 Plate: B0809 row: C column: 08
 Seq primer: M13 Reverse
 High quality sequence stop: 665
 POLYA=No.

Location/Qualifiers
 1. .665
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="NIA:B0809C08 IMAGE:3046819"
 /dev_stage="Newborn Kidney"
 /lab_host="DH10B"
 /clone_lib="NIA Mouse Newborn Kidney cDNA Library (Long 1)"
 /note="vector: pCMV-SPORT6 (Invitrogen); Site 1: SalI; Site 2: NotI; Mouse cDNA project by the Laboratory of Genetics, National Institute on Aging (NIA), Intramural Research Program, NIH (http://lgsun.grc.nia.nih.gov/cDNA). In brief, double-stranded cDNAs were synthesized with an Oligo(dT) primer [Invitrogen]; 5'-PGACTAGTCTAGATCGGAGCGCGCCCTTTT-3' from 26 ug of total RNA, treated with T4 DNA polymerase, and purified by ethanol precipitation. The cDNAs were ligated to Loe-linker L1-SalI, purified by phenol/chloroform, and separated from free linkers by Centricon 100. Then, the cDNAs were amplified by long-range high fidelity PCR using Ex Taq polymerase (Takara) with a primer Sal4-S. The products were purified by phenol/chloroform and Centricon 100. The cDNAs were digested with SalI and NotI enzymes and cloned into SalI/NotI site of pCMV-SPORT6 plasmid vector. The DH10B E. coli host was transformed with the ligation mixture by the standard chemical method. The average insert size is about 3.0 kb. The library was constructed by Yulan Piao."

FEATURES
 source

ORIGIN

Query Match 23.7%; Score 399.4; DB 14; Length 665;
 Best Local Similarity 76.8%; Pred. No. 1.4e-102;
 Matches 503; Conservative 0; Mismatches 146; Indels 6; Gaps 1;

QY 1 ATGTTTGCAGGATCTGATCTCCATGTTCTGCTGTTAATGGTCTGGTGGGAAGACA 60
 DB 17 ATATTTGTCAGGATGTTGGTGTTCGTTCTGCTGCTAATCGCCCTGGTGGGAAGTCA 76
 QY 61 GCCTGTGGGTTCTCCCTGATGTCCTTTATTGGAAGCCCTGGACCCAGACTGACCCCTGAC 120
 DB 77 GTCAATGGGCTCTCACTGATGTCCTTCATTGGCGCCCGCCAGACCCAGATTGGACCCCGAT 136
 QY 121 CAGTATGATTACAGTACAGGATTAATCAGGAAGAGAACACCCAGTAGCACACTTACC 180
 DB 137 GACTATTACTAGTATGAGCAGTCCAGCCAGAGAGACCCAGTGTACGAGACACC 196
 QY 181 CATGTGAGATCTGACTGGTACTACACTGAGGACCAAGTGTATCCATGTCAGCCCAAC 240
 DB 197 ACCCTGTGAGACCCGACTGGTACT-----ATGAAGACGATGATCCATGCGATCCAAAC 250
 QY 241 CCCTGTGACAGGGTGGGACTGCTCGTCCATGGGAGACCTTCACATGAGTGGCTG 300
 DB 251 CCCTGTGAAACAGCGGGGACTGTATCATCAGAGGGGATACCTTCAGTTGAGCTGCCCA 310
 QY 301 GCTCCTTTCTCTGGGAATAGTGTGAGAAAGTGCAGAAATACGTCAAGAGCAACCCCATGT 360
 DB 311 GCCCCCTTCTCGGGAGCGGTGCAGACTGACACAAACAAAGTGCAGAGCAACCCCATGT 370
 QY 361 GGCGGGGCCATGTCTGATACCCAGAGTCTCCCTACTACCGCTGTCTGTAACAC 420
 DB 371 GTCCATGGTGTGCTCTATTACCCAGAGACCCCTACTACCGCTGTGCTGCAATAC 430
 QY 421 CTTTACACAGGTCCTGCTCCCAAGTGGTTCTGTATGAGGCCCAAAACCCCTGCCAG 480

DB 431 CCTTACAGGGACCAGACTGCTCCAAAGTGTTCGGCATGAGGCCAAACCCCTGCCAG 490
 QY 481 AATGGGGTACTCTGCTCCCGGATAGCGGAGATCCAAAGTTTCACTGTGCTGTCCCGAC 540
 DB 491 AATGGGGAGTCTGTTCCGACACAGAGGAGATCCAGTTTACTGTGCTGTCCAGAC 550
 QY 541 CAGTTCAAGGGGAATCTGTGAATAGTTCTGATGATCTGCTATGTGGGATGGCTAC 600
 DB 551 CAGTATAAGGGGAATCTGTGAATAGTCTCGGACGACTGTATGTGCTGCTGATGGCTAC 610
 QY 601 TCTTACCGGGGAATTAAGTAGGACAGTCAACACGATCGTGGCTGCTTTACTTGA 655
 DB 611 TCTTACCGGGGAATTAAGTAGGACAGTCAACACGATCGTGGCTGCTTTACTTGA 665

RESULT 15
 BI332440
 LOCUS
 DEFINITION 791 bp mRNA linear EST 30-JUL-2001
 602380841F1 NCI_CGAP_Li9 Mus musculus cDNA clone IMAGE:5133587 5', mRNA sequence.
 ACCESSION BI332440
 VERSION BI332440.1 GI:15017097
 KEYWORDS EST.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 791)
 NIH-MGC http://mgi.nci.nih.gov/.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabbs@mail.nih.gov
 Tissue Procurement: Jeffrey E. Green, M.D.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLML1328 row: a column: 12
 High quality sequence stop: 787.

FEATURES
 Location/Qualifiers
 1. .791
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="FVB/N"
 /db_xref="taxon:10090"
 /clone="IMAGE:5133587"
 /lab_host="DH10B (T1 phage-resistant)"
 /clone_lib="NCI_CGAP_Li9"
 /note="Organ: liver; Vector: pCMV-SPORT6; Site 1: NotI; Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 1.9 kb. Constructed by Life Technologies. Note: this is a NCI_CGAP Library."

ORIGIN

Query Match 23.6%; Score 396.4; DB 12; Length 791;
 Best Local Similarity 75.6%; Pred. No. 1.1e-101;
 Matches 520; Conservative 0; Mismatches 161; Indels 7; Gaps 2;

QY 1 ATCTTCCAGGATGCTGATCTCCATGTTCTGCTGTTAATGGCTGTGGTGGGAAGACA 60
 DB 23 ATATTTGTACGAGTGTGGTGTTCCTGTCTCTGCTAATCGCCCTGTGGGAAATCA 82
 QY 61 GCCTGTGGGTTCTCCCTGATGTCCTTTATTGGAAGCCTCGACCCAGACTGGACCCCTGAC 120
 DB 83 GTCAATGGGCTCTCACTGATGTCCTTCATTGGCCCCCAGACCCAGATTGGACCCCGAT 142
 QY 121 CAGTATGATTACAGTACGAGGATTAATCAGGAGAGACACACAGTAGGACACTTACC 180
 DB 143 GACTATTACTACGCTATGAGCAGTCCAGCCAGACGAGACCCAGTGTACGCGAGCC 202

QY 181 CATGCTGAGATCCTGACTGCTACTACTAGAGACCAAGCTGATCCATGCGAGCCCAAC 240
Db |||||
QY 203 ACCCTGAGAACCCGAGCTGCTACT------ATGAAGACGATGATCCATGCGAGCCCAAC 256
Db |||||
QY 241 CCCTGTGAACACGCTGGGAGCTGCTGCTCCATGGGAGCACCTTCACATGAGCTGCTG 300
Db |||||
QY 257 CCCTGTGAACACGCTGGGAGCTGCTGCTCCATGAGGGGATACCTTCAGTTGAGCTGCCCA 316
Db |||||
QY 301 GCTCCTTTCTCTGGGAATAGTGTGAGAAAGTGCAGAAATAGTGCAGAGGACACCCATGT 360
Db |||||
QY 317 GCCCCCTCTCTGGGAGCGGCTGCGAGCTGCACAAAACAGTGCAGGACACCCATGT 376
Db |||||
QY 361 GGCCGGGGCCAAATGTCTCATTAACCCAGAGTCCCTCCTACTACCGCTGTGTGTAAACAC 420
Db |||||
QY 377 GTCCATGGTGATTGCTCTATTACCCAGAGACCCCTACTACCGCTGTGTGTAAATAC 436
Db |||||
QY 421 CCTTACAGGTCCAGCTGCTCCCAAGTGGTTCCTGTATGCGAGCCCAACCCCTGCCAG 480
Db |||||
QY 437 CCTTACAGGGACCAAGCTGCTCCAAAGTGTTCGGGATGCGAGCCCAACCCCTGCCAGA 496
Db |||||
QY 481 AATGGGGCTACCTGCTCCCGGATAGCGGAGATCCAGTTCACTGTGCTGCTCCGAC 540
Db |||||
QY 497 ATGGCCGGAGTCTGTTCCCGACACAGACGGAGATCCAGTTTACCTGTGCTGTCCAGAC 556
Db |||||
QY 541 CAGTTCAAGGGGAAATCTGTGAA-ATAGTTCTGATGACTGCTATGTTGGGATGGCTA 599
Db |||||
QY 557 CAGTATAAGGGGAAATCTGTGAACATAGTCCGGACGACTGTTATGTCGGGTGATGGCTA 616
Db |||||
QY 600 CTCCTTACCGAGGAAATCAATAGGACAGTCAACAGCATGCGTGCCTTTACTGGAATC 659
Db |||||
QY 617 CTCCTTACCGAGGAAAGTGAAGACAGTTCAACAGACCCATGCCITTTACTGGAATC 676
Db |||||
QY 660 CCACCTCCTCTTGCAGGAGAAATTACAAC 687
Db |||||
QY 677 CCAGCTCCTCTTGCAGGAGACTTTATAAC 704
Db |||||

Search completed: May 25, 2004, 10:42:53
Job time : 3061 secs

Sequence 5, Appli
Sequence 5, Appli
Sequence 64, Appli
Sequence 66, Appli
Sequence 50, Appli
Sequence 54, Appli
Sequence 56, Appli
Sequence 62, Appli
Patent No. 5200340
Sequence 10, Appli
Sequence 38, Appli
Sequence 42, Appli
Sequence 3, Appli
Sequence 25, Appli
Patent No. 5200340
Sequence 3, Appli

28 70.6 4.2 1163 2 US-08-558-269-5
29 70.6 4.2 1163 3 US-09-410-882-5
30 70.6 4.2 1170 2 US-08-811-949-64
31 70.6 4.2 1170 2 US-08-811-949-66
32 70.6 4.2 1314 2 US-08-811-949-50
33 70.6 4.2 1314 2 US-08-811-949-54
34 70.6 4.2 1314 2 US-08-811-949-62
35 70.6 4.2 1419 2 US-08-811-949-62
36 70.6 4.2 1738 6 5200340-1
37 70.6 4.2 1848 3 US-08-814-412-10
38 70.6 4.2 1974 2 US-08-811-949-38
39 70.6 4.2 2101 2 US-08-811-949-42
40 70.6 4.2 2162 1 US-08-113-512-3
41 70.6 4.2 2162 1 US-08-488-015B-3
42 70.6 4.2 2162 1 US-08-488-015B-25
43 67.8 4.0 1724 6 5200340-5
44 67.8 4.0 2497 6 5185259-2
45 66.4 3.9 2544 3 US-09-518-046-3

ALIGNMENTS

RESULT 1
US-08-148-910-14
; Sequence 14, Application US/08148910
; Patent No. 5466593
; GENERAL INFORMATION:
; APPLICANT: Takeshi SHIMOMURA et al.
; TITLE OF INVENTION: No. 5466593el Protein and Gene Encoding Said Protein
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack
; STREET: 805 Fifteenth Street, N.W., #700
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch,
; MEDIUM TYPE: 500 Kb Storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/148,910
; FILING DATE: No. 5466593ember 5, 1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Warren M. Cheek, Jr.
; REGISTRATION NUMBER: 33,367
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-8850
; TELEFAX: 202-371-8856
; TELEX:
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2033 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; ORIGINAL SOURCE:
; ORGANISM: human
; IMMEDIATE SOURCE:
; LIBRARY: Pre-made Lambda phage Library,
; LIBRARY: human liver(49, male) cdna Library (Stratagene)
US-08-148-910-14

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 25, 2004, 05:49:14 ; Search time 91.5 Seconds
(without alignments)
10207.457 Million cell updates/sec

Title: US-09-912-559-2
Perfect score: 1683
Sequence: 1 atgtttccaggatgtctga.....aaagtgaagtggcttcttaa 1683

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA.*
1: /cgn2_6/ptodata/2/ina/5A COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PCTUS COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	114.4	6.8	2033	1	US-08-148-910-14
2	114.4	6.8	2033	1	US-08-448-937A-14
3	93	5.5	970	1	US-08-148-910-3
4	93	5.5	970	1	US-08-448-937A-3
5	72.2	4.3	329	1	US-08-148-910-2
6	72.2	4.3	329	1	US-08-148-910-13
7	72.2	4.3	329	1	US-08-448-937A-2
8	72.2	4.3	329	1	US-08-448-937A-13
9	72.2	4.3	1065	1	US-08-427-640-1
10	72.2	4.3	1065	1	US-08-427-640-5
11	72.2	4.3	1068	1	US-08-427-640-3
12	72.2	4.3	1137	4	US-09-553-498-9
13	72.2	4.3	1137	4	US-09-618-868-9
14	72.2	4.3	1314	2	US-08-811-949-48
15	72.2	4.3	1314	2	US-08-883-795A-39
16	72.2	4.3	2457	6	5344773-1
17	72.2	4.3	2544	4	US-09-703-695A-3
18	72.2	4.3	7360	1	US-08-286-740-1
19	72.2	4.3	7360	5	PCT-US95-09576-1
20	70.6	4.2	1068	2	US-08-811-949-60
21	70.6	4.2	1068	1	US-08-137-116-2
22	70.6	4.2	1068	1	US-08-427-640-7
23	70.6	4.2	1068	2	US-08-811-949-44
24	70.6	4.2	1068	2	US-08-811-949-46
25	70.6	4.2	1068	2	US-08-811-949-52
26	70.6	4.2	1068	2	US-08-811-949-58
27	70.6	4.2	1068	6	5223256-3

Query Match 6.8%; Score 114.4; DB 1; Length 2033;
Best Local Similarity 49.6%; Pred. No. 3.5e-25;
Matches 605; Conservative 0; Mismatches 561; Indels 54; Gaps 10;

QY 469 AACCCCTGCCAGAAATGGGGTACTCTGCTCCCGGCAATAAGCGGAGATCCAAAGTTCACTGT 528
DB 742 AGCCCTTGGCTGAACGGGGCACCTGCCACCTGATCGTGGCCACCGGACCAACCGTGTGT 801

QY 529 GCCTGCTCCCGACAGTTTCAAGGGGAAATCTGTGAAATAGTTTCTGATGA---CTGCTAT 585
DB 802 GCCTGCTCCCGACAGCTTTCGTGACGCTCTGCAATCATCGAGCTGATGAGCGCTGCTTC 861

QY 586 GTTGGCGATGCTACTCTTACCGAGGGGAAATGAATAGGACATCAACAGCAGTCGCTGC 645
DB 862 TTGGGGAACGGCACTGGGTACCGTGGCGTGGCCAGCACCTCAGCCTCGGGCTCAGCTGC 921

QY 646 CTTTACTGGAACTCCCACTCTCTTCCAGAGAAATTTACAACATGTTTATGAGGATGCT 705
DB 922 CTGGCTTGGAACTCCGATCTGCTTACAGAGACTGCAGCTGGACTCCGTGGCGCGCG 981

QY 706 GAAACCCATGGATTTGGGGACACAATTTCTGCAGAAACCCAGATCGGACGAAAAGCCC 765
DB 982 GCCCTGTGGCTGGCGCCCATGCTCTGCGGAATCCGGACAATGACGAGAGGCC 1041

QY 766 TGGTGTCTTATTAAGTTTCAATGACAGAGTGAATGGGAATCTGATGCTCAGCC 825
DB 1042 TGGTGTCTACGT---GGTGAAGGACAGCGCGCTCTCTGGAGTACTCCGCCCTGGAGGCC 1098

QY 826 TGCTCAGCCAGGAGCTTGCCTACCCAGAGGAAAGCCCACTGAGCCATCAACCAAGCTT 885
DB 1099 TGGCAATCCCT-----CACAGAGTCCAACTGTACCGGATCTCTGGCGACCTG 1149

QY 886 CCGGGTGTGACTCTGTGAAAGACTGAGATAGCAGAGAGGAGATCAAGAGATCTAT 945
DB 1150 CTTGAGCGAGCTCTCCCGGGCGCGCAGGCTCTGCGGAGAGGACAAAGAGAGACGCTT 1209

QY 946 GGAGCTTTAAGACAGCGCGGGCAAGCACCCATCGGAGGCGTCCCTCCAGTCTCTGCTG 1005
DB 1210 CTGGCGCACGATCATCTCGCGGCTCTCTCTGCTGCTCGCGCTCGCACCCCTG----- 1262

QY 1006 CTTGACATCTCCATCCCGGCGCACTCTGTGTGGGGCGCTGATCCACCCCTGC 1065
DB 1263 --GCTGGCGGCATCTACATCGGGGACAGCTCTTGGCGGGAGCGCTGGTCCACACCTGC 1320

QY 1066 TGGTGTCTCACTGTGCGCACTGACCGACACATAAAACAGACATCTAA-----AGGTG 1119
DB 1321 TGGGTGTGTGGCGCGCCCACTGCTCTCCACAGCCCGCCCGGACAGCGTCTCGTG 1380

QY 1120 GTGCTAGGGACCGAGCCTGAAGAAAGAAATTTATGAGCAGAGCTTTAGGTCAG 1179
DB 1381 GTGCTGGCCAGCACTTTCAACCGCACGACGAGCGAGCTTGGCGACGCTTGGCATCGAG 1440

QY 1180 AAGATATTCAAGTACAGCACTACAAATGAAGAGATGAGATTCGCCCAATGATATGCA 1239
DB 1441 AAGTACATCCGTACACCTGTACTCGGTGTTCAACCCAG---CGACACAGCTCTGTC 1497

QY 1240 TTGCTCAAGTTAAAGCCAGTGGATGGTCACTGTGCTTAGAATCCAAATAGTGAAGACT 1299
DB 1498 CTGATCCGGCTGAAGAAGAGGGAGCGCTGTGCCACACGCTGCGAGTTCTGTCAGGCC 1557

QY 1300 GTGCTGCTGCTGATG-----GPTCTTTCCTCTGGGAGTGGAGTGCACATCTCTGGC 1353
DB 1558 ATCTGCTGCTGCGAGCCCGGAGCACCTTCCCGGAGGACACAGTGCAGATTGCGGGC 1617

QY 1354 TGGGG-----TGTTACAGAAACAGGAAAGGGTCCCGCCAGCTCTCGATGCCAAAGTC 1407
DB 1618 TGGGGCCACTTGGATGAGAACGTGAGCGGCTACTCCAGCTCCCTGCGGGAGGCGCTGTC 1677

QY 1408 AAGCTGATTGCCAACATTTGTGCACTCCCGCACTCTATGACCATGATGATGATGAC 1467
DB 1678 CCCCCTGGTGGCGAGCAACAAAGTGTGAGAGCCCTGAGTCTTACGGGCGCAGATCAGCCCC 1737

QY 1468 AGTATGATCTGTGAGGAAATCTTCAGAAACCTTGGGCAAGACACCTGCCAGGGTGA 1527

DB 1738 AACATGCTCTGTGCGGCTACTT---CGACTGAAGTCCGACGCTGCCAGGGGAGCTCA 1794

QY 1528 GGAGGCCCTGACTCTGTGAGAGAGCGGACCTACTTACTGCTCTATGGGATAGTGAAGCTGG 1587

DB 1795 GGGGGGCCCCCTGGCTCTGCGAGAGAACGGGTGCTTACTCTTACGGCATCAACAGCTGG 1854

QY 1588 GGCCTGGAGTGTAG-----AAGAGCCAGGGGTCTTACACCAAGTTACAAATTCCTG 1641

DB 1855 GGTGACGCTCGGGCGGCTCCACAGCGGGGTCTTACACCGCGTGGCGCAACTATGTG 1914

QY 1642 AATTGGATCAAGCCACCAT 1661

DB 1915 GACTGGATCAACGACCGAT 1934

RESULT 2
US-08-448-937A-14
; Sequence 14, Application US/08448937A
; Patent No. 5677154
; GENERAL INFORMATION:
; APPLICANT: Takeshi SHIMOMURA et al.
; TITLE OF INVENTION: No. 5677154e1 Protein and Gene Encoding Said Protein
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack
; STREET: 805 Fifteenth Street, N.W., #700
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch,
; MEDIUM TYPE: 500 Kb Storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/448,937A
; FILING DATE: May 24, 1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/148,910
; FILING DATE: No. 5677154ember 5, 1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Warren M. Cheek, Jr.
; REGISTRATION NUMBER: 33,367
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-8850
; TELEFAX: 202-371-8856
; TELEX:
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2033 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; ORIGINAL SOURCE:
; ORGANISM: human
; IMMEDIATE SOURCE:
; LIBRARY: Pre-made Lambda phage Library,
; LIBRARY: human liver(49, male) cDNA Library (Stratagene)
; US-08-448-937A-14

Query Match 6.8%; Score 114.4; DB 1; Length 2033;
Best Local Similarity 49.6%; Pred. No. 3.5e-25;
Matches 605; Conservative 0; Mismatches 561; Indels 54; Gaps 10;

QY 469 AACCCCTGCCAGAAATGGGGTACTCTGCTCCCGGCAATAAGCGGAGATCCAAAGTTCACTGT 528
DB 742 AGCCCTTGGCTGAACGGGGCACCTGCCACCTGATCGTGGCCACCGGACCAACCGTGTGT 801

QY 529 GCCTGTCCGACAGTTCAAGGGAAATTCGTGAAATAGTCTCTGATGA---CTGCTAT 585
 Db 802 CCCTGCCACACAGGCTTCGCTGGAGCGCTCTGCAATCGAGCTGATGAGCGCTGCTTC 861
 QY 586 GTTGGCGATGCTACTCTTACCGAGGGAATGAATAGGACAGTCAACCGACATGCGTGC 645
 Db 862 TTGGGGAACGGCAGCTGGGTACCGTGGGTGGTGGCCAGCACCTCAGCCTCGGCGCTAGCTGC 921
 QY 646 CTTTACTGGAATCCCACTCTCTTTCAGAGAAATTAACAATGTTTATGGAGATGCT 705
 Db 922 CTGGCTGGAACTCGATCTGCTTACAGAGAGCTGACGTGGACTCGTGGGCGCGCG 981
 QY 706 GAAACCCATGGATTTGGGAAACAAATTTTCAGAAACCCAGATGCGGACGAAAGCC 765
 Db 982 GCCCTGTGGGCTGGGCGCCCATGCTTACTGCGGAATCCGGCAATGACGAGAGGCC 1041
 QY 766 TGGTGTCTTATTAAGTTACCAATACAGAGTGAATGGGAATCTGATGCTCTAGCC 825
 Db 1042 TGGTGTCTAGT---GGTGAAGAGACAGCGCTCTCTGGGAGTACTCGCGCTGGAGGCC 1098
 QY 826 TGCTCAGCCAGGAGTTGGCTTACCCAGAGGAAAGCCCACTGAGCCATCAACCAAGCTT 885
 Db 1099 TGGGAATCCCT-----CACAGAGTCCAACTGTCAACGGATCTCTGGCGACCTG 1149
 QY 886 CCGGGTTGACTCTGTGGAAGACTGAGATAGCAGAGAGAGATCAAGAGATCTAT 945
 Db 1150 CTTGAGCAGGCTCCCGGGCGCGCAGGCTTCTGGCAGGAGGCAAGAGAGAGCTTC 1209
 QY 946 GGAGGCTTAAAGACAGCGCGGCAAGCACCATGCGAGGGCTCCCTCCAGTCTCTCGCTG 1005
 Db 1210 CTGGCGCACATATCATCGGGCTCTCTCTCGTGGCGGCTCGACCCCTG----- 1262
 QY 1006 CTTCTGACCATCTCATGCCCGCAGGGCCACTTCTGTGTGGGGGCTGATCCACCCCTGC 1065
 Db 1263 --GCTGGCGCCATCTATATCGGGGACAGCTTCTCGCGGAGCGCTGTCTCCACCTGC 1320
 QY 1066 TGGGTGCTCACTGCTGCCACTGACCGACATATAAAGACAGACATCTAA-----AGGTG 1119
 Db 1321 TGGTGTGTGCGCGCCCACTGCTCTCCACAGCCCGCCAGGACAGGCTCTCGTG 1380
 QY 1120 GTGTAGGGGACAGGACTGAGAAAGAAAGAAATTTCAATGAGCAGAGCTTTAGGGTGAG 1179
 Db 1381 GTGTGGGCGAGCACTTCTTCAACCGCAGCAGGAGCTGACGACACCTTCGGCATCAG 1440
 QY 1180 AAGATATTCAAGTACAGCCACTTACATGAAGAGATGAGATTCGCCCAATGATATTGCA 1239
 Db 1441 AAGTACATCCGCTACACCTGCTACTCGGTGTTCAACCCAG---CGACACAGCTCTGTC 1497
 QY 1240 TTGCTCAAGTTAAGCCAGTGGATGCTACTGTCTTAGAATCCAAATACGTGAAGACT 1299
 Db 1498 CTGATCCGCTGAAGAAAGAGGAGCCGCTGTGCCACACGCTCGAGTTCTGTGAGGCC 1557
 QY 1300 GTGTGTTGCTGATG-----GGTCTTCTCTGAGTGGAGTGCCACATCTCTGCG 1353
 Db 1558 ATCTGCTCCGAGCCCGGAGCACTTCCCGGAGGACACAGTGGCCAGATTTCGGGCG 1617
 QY 1354 TGGGG-----TGTTACAGAAACAGGAAAGGTCCCGCAGCTCTCTGATGCCAAGTC 1407
 Db 1618 TGGGGCACTTGGATGAGAACGTCGAGCGCTACTCCAGCTCTCTGCGGAGGCCCTGTGTC 1677
 QY 1408 AAGCTGATGCAACACTTGTGCAACTCCGCAACTCTATGACCATGATGATGATGAC 1467
 Db 1678 CCCTGTGTCGCGACCAAGTGCAGACGCTTGTAGGTCTACGGGCGGACATCAGCCCG 1737
 QY 1468 AGTATGATCTGTGAGGAATCTTCAGAAACCTTGGGCAAGACACCTGCCAGGGTGAATCT 1527
 Db 1738 AACATGCTGTGTCGCGCTACTT---CGACTGCAAGTCCGAGCGCTGCCAGGGGACTCA 1794
 QY 1528 GAGAGGCCCTGACTGTGAGAGAGGACGACCTACTAGTCTATGGATATGAGCTGG 1587
 Db 1795 GGGGGGCTTGGCTCGAGAGAGAGCGGCTTACCTCTACGGCATCATCAGCTGG 1854

QY 1588 GGCCTGGAGTGTGAG-----AAGAGGCCAGGGGCTTACACCAAGTTACCAATTCCTG 1641
 Db 1855 GGTGAGCGGTGGGGGCTCCACAAGCGGGGCTTACACCCGCGTGGCCAACTATGTG 1914
 QY 1642 AATTGGATCCARAGCCACCAT 1661
 Db 1915 GACTGATCACACCGGAT 1934
 RESULT 3
 US-08-148-910-3
 ; Sequence 3, Application US/08148910
 ; Patent No. 546593
 ; GENERAL INFORMATION:
 ; APPLICANT: Takeshi SHIMOMURA et al.
 ; TITLE OF INVENTION: No. 546593el Protein and Gene Encoding Said Protein
 ; NUMBER OF SEQUENCES: 14
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Wenderoth, Lind & Ponack
 ; STREET: 805 Fifteenth Street, N.W., #700
 ; CITY: Washington
 ; STATE: D.C.
 ; COUNTRY: U.S.A.
 ; ZIP: 20005
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette, 5.25 inch,
 ; MEDIUM TYPE: 500 Kb Storage
 ; COMPUTER: IBM Compatible
 ; OPERATING SYSTEM: MS-DOS
 ; SOFTWARE: Wordperfect
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/148,910
 ; FILING DATE: No. 546593ember 5, 1993
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER:
 ; FILING DATE:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Warren M. Cheek, Jr.
 ; REGISTRATION NUMBER: 33,367
 ; REFERENCE/DOCKET NUMBER:
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 202-371-8850
 ; TELEFAX: 202-371-8856
 ; TELEX:
 ; INFORMATION FOR SEQ ID NO: 3:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 970 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: cDNA
 ; ORIGINAL SOURCE:
 ; ORGANISM: human
 ; IMMEDIATE SOURCE:
 ; LIBRARY: Pre-made Lambda phage library, human liver (49, male)
 ; LIBRARY: cDNA Library (Stratagene)
 ; US-08-148-910-3
 Query Match 5.5%; Score 93; DB 1; Length 970;
 Best Local Similarity 52.7%; Pred. No. 1.1e-18;
 Matches 357; Conservative 0; Mismatches 290; Indels 30; Gaps 6;
 QY 1009 CTGACCATCTCCATGCCCCCAGGGCCACTTCTGTGGTGGGCGCTGATCCACCCCTGCTGG 1068
 Db 201 CTGGCCGCCATCTACATCGGGGACAGCTTCTCGCCGGGAGCCTTGTTCACACTGCTGG 260
 QY 1069 GTGCTCACTGTGCGCCACTGCACCGACATATAAACACAGACA-----TCTAAAGGTGGTG 1122
 Db 261 GTGGTGTGGCGCGCCACTGCTTCTCCACAGCCCCCAGGAGACAGCTCTCCGTGGTG 320
 QY 1123 CTAGGGACACGAGACCTGAAGAAAGAAATTTTCATGAGCAGAGCTTAGGGTGCAGAG 1182

Db 321 CTGGGCGCAGCATTCTTCAACCGCAGCAGGACGCGACGCTTCCGGCATCGAGAAG 380
 QY 1183 ATATTCAAGTACAGCCACTTACAAATGAAGAAGATGAGATTCGCCCAATGATATTCGATTG 1242
 Db 381 TACATCCCGTACACCTGTACTCGGTGTTCACCCCA---CGCACCACGACCTCGTCTCG 437
 QY 1243 CTCAAAGTTAAAGCCAGTGTGCTACTGCTCTAGATCCAAATACGCTGAGACTGTG 1302
 Db 438 ATCCGGCTGAAGAAGAGGGACCGCTGTGCCACACGCTCGAGTTCGTGACGCCATC 497
 QY 1303 TGCCTGCTGAG---GTCCTTCCCTCTGGAGTGTGAGTCCCAATCTCTGCTGG 1356
 Db 498 TGCCTGCGCGACCGCGCAGCAGCCTTCCCGCAGGACACAAAGTCCAGATTGGGGCTGG 557
 QY 1357 GG-----TGTTACAGAAACAGGAAAGGTCGCGCAGCTCTCTGGATGCCAAAGTCAAG 1410
 Db 558 GCCCACTTGGATGAGAAAGTGTGAGCGGTACTCTCGAGCTCCCTGCGGAGGCCCTGTTCCC 617
 QY 1411 CTGATTGCCAACACTTTGTGCACTCCGCCCACTCTATGACCAATGATGATGACAGT 1470
 Db 618 CTGGTGGCGGACCAAGTGTGAGCAGCAGCCTGTAGGTCTACGGCGCGCATCAGCCCCAAC 677
 QY 1471 ATGATCTGTGCAAGGAAATCTTTCAGAAACCTGGGCAAGACCTGCCAGGGTGAATCTGGA 1530
 Db 678 ATGCTCTGTGCGGCTACTT---CGACTGCAAGTCCGAGCGCTGCCAGGGGACTCAGGG 734
 QY 1531 GCCCCCTGACTGTGAGAGGACGGCACCTACTACTGTCTATGGGATGATGAGCTGGGGC 1590
 Db 735 GGGCCCCCTGGCTGCGAAGAACGGCTGTACCTTACCTTACCGGCATCATCAGCTGGGGT 794
 QY 1591 CTGGAGTGTGAG-----AAGAGGCCAGGGGTCTACACCCCAAGTTACCAATTCCTGAAT 1644
 Db 795 GACGGCTGGGCGGCTCCACAAAGCGGGGTCTACACCGCGTGGCCAACTATGTGGAC 854
 QY 1645 TGGATCAAGCCACCAT 1661
 Db 855 TGGATCAAGCCGGAT 871

RESULT 4

US-08-937A-3
 ; Sequence 3, Application US/08448937A
 ; Patent No. 567164

GENERAL INFORMATION:

APPLICANT: Takeshi SHIMOMURA et al.
 TITLE OF INVENTION: No. 5677164el Protein and Gene Encoding Said Protein
 NUMBER OF SEQUENCES: 14
 CORRESPONDENCE ADDRESS:
 ADDRESSES: Wenderoth, Lind & Penack
 STREET: 805 Fifteenth Street, N.W., #700
 CITY: Washington
 STATE: D.C.
 COUNTRY: U.S.A.
 ZIP: 20005

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 5.25 inch,
 MEDIUM TYPE: 500 Kb Storage
 COMPUTER: IBM Compatible
 OPERATING SYSTEM: MS-DOS
 SOFTWARE: Wordperfect

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/448,937A
 FILING DATE: May 24, 1995
 CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/148,910
 FILING DATE: No. 5677164ember 5, 1993
 ATTORNEY/AGENT INFORMATION:
 NAME: Warren M. Cheek, Jr.
 REGISTRATION NUMBER: 33,367
 REFERENCE/DOCKET NUMBER:
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 202-371-8850

TELEFAX: 202-371-8856
 TELEX:
 INFORMATION FOR SEQ ID NO: 3:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 970 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: cDNA
 ORIGINAL SOURCE:
 ORGANISM: human
 IMMEDIATE SOURCE:
 LIBRARY: Pre-made Lambda phage Library, human liver (49, male)
 LIBRARY: cDNA Library (Stratagene)
 US-08-448-937A-3

Query Match 5.5%; Score 93; DB 1; Length 970;

Best Local Similarity 52.7%; Pred. No. 1.1e-18;

Matches 357; Conservative 0; Mismatches 290; Indels 30; Gaps 6;

QY 1009 CTGACCATCTCCATGTCGCCAGGGCCACTTCTGTGTGGGCGCTGATCCACCCCTCTGCTG 1068
 Db 201 CTGGCCGCCATCTACATCGGGGACAGCTTCTGCGCGGGAGCTGTCCACACCTGCTGG 260
 QY 1069 GTGCTCACTGTCGCCACTGCACGGACATATAAAACACAGACA-----TCTAAGGTGGT 1122
 Db 261 GTGCTGTGGCGCGCCACTCTTCTCCACAGCCCCCAGGGACAGCGTCTCCGTTGGT 320
 QY 1123 CTAGGGACACGAGGACCTTGAAGAAAGAAATTTTCATGACGAGAGCTTTAGGGTGCAGAA 1182
 Db 321 CTGGCCGAGCCTCTTCTCAACCCGACGACGAGCTGACGACGAGCTTCGGCATCGAGAAG 380
 QY 1183 ATATTCAAGTACAGCCACTTACAAATGAAGAAGATGAGATTCGCCCAATGATGATGATG 1242
 Db 381 TACATCCCGTACACCTGTACTCGGTGTTCACACCCCA---GGACACACCTCTGCTCTG 437
 QY 1243 CTCAAGTTAAAGCCAGTGGTGTGCTGTGCTCTAGATTCCTAGATTCCTAGATTCCTAG 1302
 Db 438 ATCCGGCTGAAGAAAGAGGGGACCGCTGTGCCACACGCTCGCAGTTCGTGAGCCCATC 497
 QY 1303 TGCCTGCTGATG-----GGTCTTCTCTCTGGAGTGTGAGTGCACATCTCTGCTGG 1356
 Db 498 TGCCTGCGCGAGCCCGGACGACCTTCCCGCGAGGACACAAGTCCAGATTGCGGGCTGG 557
 QY 1357 GG-----TGTTACAGAAACAGGAAAGGGTCCCGCCAGCTCTCTGATGCCAAAGTCAAG 1410
 Db 558 GGCCACTTGGATGAGAAAGTGTGAGCGGTACTCTCGAGCTCCCTGCGGAGGCCCTGTTCCC 617
 QY 1411 CTGATTGCCACACTTGTGCAACTCCGCCACTCTATGACCAATGATGATGATGATGATGATG 1470
 Db 618 CTGGTCCCGACCAACAAGTGTGAGCAGCCTGTAGGTCTACGGCGCGCATCAGCCCCAAC 677
 QY 1471 ATGATCTGTGAGGAAATCTTTCAGAAACCTGGGCAAGACACCTGCCAGGGTGAATCTGGA 1530
 Db 678 ATGCTCTGTGCGGCTACTT---CGACTGCAAGTCCGAGCGCTGCCAGGGGACTCAGGG 734
 QY 1531 GGCCCTGACCTGTGAGAGGAGCGCACCTACTACTGTCTATGGGATGATGAGCTGGGGC 1590
 Db 735 GGGCCCTGCGCTGCGAAGAAAGCGGCTGTGCTTACTCTACGGCATCATCAGCTGGGGT 794
 QY 1591 CTGGAGTGTGAG-----AAGAGGCCAGGGGTCTACACCCCAAGTTACCAATTCCTGAAT 1644
 Db 795 GACGGCTGGGCGGCTCCACAAAGCGGGGTCTACACCGCGTGGCCAACTATGTGGAC 854
 QY 1645 TGGATCAAGCCACCAT 1661
 Db 855 TGGATCAAGCCGGAT 871

RESULT 5

US-08-148-910-2

; Sequence 2, Application US/08148910

; Patent No. 5466593

Patent No. 5466593
GENERAL INFORMATION:
APPLICANT: Takeshi SHIMOMURA et al.
TITLE OF INVENTION: No. 5466593el Protein and Gene Encoding Said Protein
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Wenderoth, Lind & Ponack
STREET: 805 Fifteenth Street, N.W., #700
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 5.25 inch,
MEDIUM TYPE: 500 Kb Storage
COMPUTER: IBM Compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: Wordperfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/148,910
FILING DATE: No. 5466593ember 5, 1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Warren M. Cheek, Jr.
REGISTRATION NUMBER: 33,367
REFERENCE/DOCKET NUMBER:
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-371-8850
TELEFAX: 202-371-8856
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 329 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
ORIGINAL SOURCE:
ORGANISM: human
IMMEDIATE SOURCE:
LIBRARY: Quick-cloneTM human liver cDNA (Clonetechn)
US-08-148-910-2

Query Match 4.3%; Score 72.2; DB 1; Length 329;
Best Local Similarity 59.3%; Pred. No. 1.7e-12;
Matches 163; Conservative 0; Mismatches 103; Indels 9; Gaps 2;
QY 1381 TCCCGCAGCTCTGGATGCCAAAGTCAAGCTGATTGCCACACTTTGTGCAACTCCGCG 1440
DB 52 TCCAGCTCCCTGCGGAGGCCCTGTCCCTGCTGCGCGGCTACTT---CGACTGC 168
QY 1441 CAACTCTATGACCATGATTGATGACAGTATGATCTGTGAGGAAATCTTCAGAAACCT 1500
DB 112 GAGTCTACGGCGCCGACATCAGCCCAACATGCTGTGCGGCTACTT---CGACTGC 168
QY 1501 GGGCAAGACACTGCCAGGCTGACTTGGAGGCCCTGACCTGTGAGAGGACGGCACC 1560
DB 169 AAGTCCGACCTGCCAGGGGACTCAGGGGGGCCCTGGCTGGAGAGAACGGCGTG 228
QY 1561 TACTAGCTCTATGGATAGTGTGAGCTGGGGCCCTGGAGTGTGAG-----AAGAGGCCAGGG 1614
DB 229 GCTTACTCTACGGCATCATCAGCTGGGTGACGGTGGCGGCGCTCCACAGCCGGGG 288
QY 1615 GTCTACACCAAGTTACCAATTCCTGAATTGGAT 1649
DB 289 GTCTACACCGCGTGGCCCAACTATGTGGACTGGAT 323

RESULT 6
US-08-148-910-13
; Sequence 13, Application US/08148910

Patent No. 5466593
GENERAL INFORMATION:
APPLICANT: Takeshi SHIMOMURA et al.
TITLE OF INVENTION: No. 5466593el Protein and Gene Encoding Said Protein
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Wenderoth, Lind & Ponack
STREET: 805 Fifteenth Street, N.W., #700
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 5.25 inch,
MEDIUM TYPE: 500 Kb Storage
COMPUTER: IBM Compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: Wordperfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/148,910
FILING DATE: No. 5466593ember 5, 1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Warren M. Cheek, Jr.
REGISTRATION NUMBER: 33,367
REFERENCE/DOCKET NUMBER:
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-371-8850
TELEFAX: 202-371-8856
TELEX:
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 329 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
ORIGINAL SOURCE:
ORGANISM: human
IMMEDIATE SOURCE:
LIBRARY: Quick-cloneTM human liver cDNA (Clonetechn)
US-08-148-910-13

Query Match 4.3%; Score 72.2; DB 1; Length 329;
Best Local Similarity 59.3%; Pred. No. 1.7e-12;
Matches 163; Conservative 0; Mismatches 103; Indels 9; Gaps 2;
QY 1381 TCCCGCAGCTCTGGATGCCAAAGTCAAGCTGATTGCCACACTTTGTGCAACTCCGCG 1440
DB 52 TCCAGCTCCCTGCGGAGGCCCTGTCCCTGCTGCGCGGCTACTT---CGACTGC 168
QY 1441 CAACTCTATGACCATGATTGATGACAGTATGATCTGTGAGGAAATCTTCAGAAACCT 1500
DB 112 GAGTCTACGGCGCCGACATCAGCCCAACATGCTGTGCGGCTACTT---CGACTGC 168
QY 1501 GGGCAAGACACTGCCAGGCTGACTTGGAGGCCCTGACCTGTGAGAGGACGGCACC 1560
DB 169 AAGTCCGACCTGCCAGGGGACTCAGGGGGGCCCTGGCTGGAGAGAACGGCGTG 228
QY 1561 TACTAGCTCTATGGATAGTGTGAGCTGGGGCCCTGGAGTGTGAG-----AAGAGGCCAGGG 1614
DB 229 GCTTACTCTACGGCATCATCAGCTGGGTGACGGTGGCGGCGCTCCACAGCCGGGG 288
QY 1615 GTCTACACCAAGTTACCAATTCCTGAATTGGAT 1649
DB 289 GTCTACACCGCGTGGCCCAACTATGTGGACTGGAT 323

RESULT 7
US-08-448-937A-2

Sequence 2, Application US/08448937A
Patent No. 5677164
GENERAL INFORMATION:
APPLICANT: Takeshi SHIMOMURA et al.
TITLE OF INVENTION: No. 5677164el Protein and Gene Encoding Said Protein
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Wenderoth, Lind & Ponack
STREET: 805 Fifteenth Street, N.W., #700
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 5.25 inch,
MEDIUM TYPE: 500 Kb Storage
COMPUTER: IBM Compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: Wordperfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/448,937A
FILING DATE: May 24, 1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/148,910
FILING DATE: No. 5677164ember 5, 1993
ATTORNEY/AGENT INFORMATION:
NAME: Warren M. Cheek, Jr.
REGISTRATION NUMBER: 33,367
REFERENCE/DOCKET NUMBER:
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-371-8850
TELEFAX: 202-371-8856
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 329 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
ORGANISM: human
IMMEDIATE SOURCE:
LIBRARY: Quick-cloneTM human liver cDNA (Clontech)
US-08-448-937A-2

Query Match 4.3%; Score 72.2; DB 1; Length 329;
Best Local Similarity 59.3%; Pred. No. 1.7e-12;
Matches 163; Conservative 0; Mismatches 103; Indels 9; Gaps 2;
QY 1381 TCCCGCAGCTCCTGGATGCCAAAGTCAAGCTGATTGCCAACACTTTGTGCAAACTCCCGC 1440
Db 52 TCCAGCTCCCTGCGGGAGGCCCTGCTCCCTGCTGCGCGACCAAGTCAGCAGCCCT 111
QY 1441 CAACCTATGACCATGATTGATGACAGTATGATCTGTGAGGAAATCTTCAGAACTT 1500
Db 112 GAGGTCTACGGCGCGGCACATCAGCCCCAACATGCTCTGTGCGGCTACTT---CGACTGC 168
QY 1501 GGGCAAGACACTGCGAGGCTGACTCTGAGGCCCCCTGACTGTGAGAGCGGCAC 1560
Db 169 AAGTCCGAGCGCTGCCAGGGGACTCAGGGGGCCCCCTGGCTCGAGAGAACGGCGTG 228
QY 1561 TACTACGCTCTATGGGATAGTGAAGTGGGCTGTGAGTGTGAG-----AAGAGCCAGGG 1614
Db 229 GCTTACCTCTACGGCATCATCAGTGGGGTACGCGCTGCGGGCGGCTCCCAAGCCGGG 288
QY 1615 GTCTACACCCAGTTACCAAAATCTCTGAATTGGAT 1649
Db 289 GTCTACACCCGGTGGCCAACTATGTGACTGGAT 323

RESULT 8

US-08-448-937A-13
Sequence 13, Application US/08448937A
Patent No. 5677164
GENERAL INFORMATION:
APPLICANT: Takeshi SHIMOMURA et al.
TITLE OF INVENTION: No. 5677164el Protein and Gene Encoding Said Protein
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Wenderoth, Lind & Ponack
STREET: 805 Fifteenth Street, N.W., #700
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 5.25 inch,
MEDIUM TYPE: 500 Kb Storage
COMPUTER: IBM Compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: Wordperfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/448,937A
FILING DATE: May 24, 1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/148,910
FILING DATE: No. 5677164ember 5, 1993
ATTORNEY/AGENT INFORMATION:
NAME: Warren M. Cheek, Jr.
REGISTRATION NUMBER: 33,367
REFERENCE/DOCKET NUMBER:
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-371-8850
TELEFAX: 202-371-8856
TELEX:
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 329 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
ORGANISM: human
IMMEDIATE SOURCE:
LIBRARY: Quick-cloneTM human liver cDNA (Clontech)
US-08-448-937A-13

Query Match 4.3%; Score 72.2; DB 1; Length 329;
Best Local Similarity 59.3%; Pred. No. 1.7e-12;
Matches 163; Conservative 0; Mismatches 103; Indels 9; Gaps 2;
QY 1381 TCCCGCAGCTCCTGGATGCCAAAGTCAAGCTGATTGCCAACACTTTGTGCAAACTCCCGC 1440
Db 52 TCCAGCTCCCTGCGGGAGGCCCTGCTCCCTGCTGCGCGACCAAGTCAGCAGCCCT 111
QY 1441 CAACCTATGACCATGATTGATGACAGTATGATCTGTGAGGAAATCTTCAGAACTT 1500
Db 112 GAGGTCTACGGCGCGGCACATCAGCCCCAACATGCTCTGTGCGGCTACTT---CGACTGC 168
QY 1501 GGGCAAGACACTGCGAGGCTGACTCTGAGGCCCCCTGACTGTGAGAGCGGCAC 1560
Db 169 AAGTCCGAGCGCTGCCAGGGGACTCAGGGGGCCCCCTGGCTCGAGAGAACGGCGTG 228
QY 1561 TACTACGCTCTATGGGATAGTGAAGTGGGCTGTGAGTGTGAG-----AAGAGCCAGGG 1614
Db 229 GCTTACCTCTACGGCATCATCAGTGGGGTACGCGCTGCGGGCGGCTCCCAAGCCGGG 288
QY 1615 GTCTACACCCAGTTACCAAAATCTCTGAATTGGAT 1649
Db 289 GTCTACACCCGGTGGCCAACTATGTGACTGGAT 323

RESULT 9
US-08-427-640-1
; Sequence 1, Application US/08427640
; Patent No. 5658788
; GENERAL INFORMATION:
; APPLICANT: Bers et al.
; TITLE OF INVENTION: Compounds and Methods for Treatment of Thromboembolic Disorder
; NUMBER OF SEQUENCES: 28
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Eli Lilly and Company
; STREET: Lilly Corporate Center
; CITY: Indianapolis
; STATE: IN.
; COUNTRY: U.S.A.
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.0 Mb storage
; COMPUTER: Macintosh
; OPERATING SYSTEM: Macintosh
; SOFTWARE: Microsoft Word
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/427,640
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/689,410
; FILING DATE: 22 APRIL 1991
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1065 base pairs
; TYPE: Nucleic acid
; STRANDEDNESS: Double
; TOPOLOGY: Linear
; MOLECULE TYPE: DNA
US-08-427-640-1

Query Match 4.3%; Score 72.2; DB 1; Length 1065;
Best Local Similarity 50.5%; Pred. No. 3,7e-12;
Matches 340; Conservative 0; Mismatches 288; Indels 45; Gaps 5;

QY	1028	AGGCGCACCTTCTGTGGTGGGGCGTGATCCACCCCTGCTGGGTGCTCAGTCTGCCCACT	1087
DB	392	AGCGGTTCCTGTGCGGGGGGATACTATCAGTCTCTGTGGATCTCTTGGCGCCCACT	451
QY	1088	GCACCGA-----CATAAAACACAGACATCTAAAGTGGTCTAGGGGACACGAGACCTGA	1141
DB	452	GCCTCCAGGAGAGGTTTCCGCCCCACACCTGACGGTGATCTTGGGCGAGAACATACCGGG	511
QY	1142	AGAAAGAAATTCATGACGAGAGCTTTAGGGTGCAGAGATATTCAGTACAGCCACT	1201
DB	512	TGTFCCCTGCGGAGGAGGAGAGAAATTTGAAGTCGAAAAATACATTTGCCATAAGGAAT	571
QY	1202	ACAATGAAGAGATGAGATTTCCCCACAAATGATATTGCATTGCTCAAAGTTAAAGCCAGTGG	1261
DB	572	TCGAT-----GATGACACITACGACATGACATTGGCTGCTGCAGCTGAAATCGGATT	625
QY	1262	ATGGTCACTGTGCTCTAAGATCCAAATACGTGAAGACTGTGTGCTTGCCTGATGGGTCCT	1321
DB	626	CGTCCCGCTGTGCCCGAGGAGACGCGTGGTCCGCACTGTGTGCTTTCCCGCGCGGACC	685
QY	1322	TTCCCT-----CTGGGAGTCAGTSCCACTCTCTGGCTGGGGTGTTACAGAA	1369
DB	686	TGCAGCTGCCGAGCTGACGGAGTGTGAGCTCTCCGGCTACGGCAAGCATGAGGCTTGT	745
QY	1370	CAGGAAAGGGTCCCGCCAGCTCCTGGATGCAAAAGTCAAGCTGATTCACCAACTTTGT	1429
DB	746	CTCCTTCTATTTCGGAGCGGCTGAAGAGGCTCATGTTCAGACTGTACCCATCCAGCGCT	805
QY	1430	GCACTCCCGCCACTCTATGACCACATGATTGATGACAGTATGATCTGTGCAGGAATC	1489
DB	806	GCACTCACAAACATTTACTTTAACAGAACAGTTCACCGCAACATGCTGTGTGCTCGAGAC	865
QY	1490	TTCAGAAACCTGG-----GCAAGACACCTGCCAGGGTGACTCTCGAGGCC	1534

Db	866	CTCGGAGCGCGGGCCCCCAGGCAAACTTGCACGACGCTGCCACGGCGGATTCGGGAGGCC	925
Qy	1535	CCCTGACCTGTGTGAGAAGGACGCGACCTACTACTGTCTATGGGATAGTACTGGGCGCTGG	1594
Db	926	CCCTGTGTGTCTGAACGATGCGCGCATGACATTTGGTGGGCATCATCAGCTGGGCGCTGG	985
Qy	1595	AGTGTGAGAAGG-----CCAGGGGTCTACACCCCAAGTTACCAAAATTCCTGAATTGGA	1648
Db	986	GCTGTGGACAGAAGGATGCCCGGGTGTATACCAAGTTACCAACTACCTAGACTGGA	1045
Qy	1649	TCAAAGCCACCAT	1661
Db	1046	TTCGTGACAACAT	1058
<p>RESULT 10 US-08-427-640-5 ; Sequence 5, Application US/08427640 ; Patent No. 5658788 ; GENERAL INFORMATION: ; APPLICANT: Berg et al. ; TITLE OF INVENTION: Compounds and Methods for Treatment of Thromboembolic Disorder ; NUMBER OF SEQUENCES: 28 ; CORRESPONDENCE ADDRESS: ; ADDRESSEE: Eli Lilly and Company ; STREET: Lilly Corporate Center ; CITY: Indianapolis ; STATE: IN ; COUNTRY: U.S.A. ; ZIP: 46285 ; COMPUTER READABLE FORM: ; MEDIUM TYPE: Diskette, 3.50 inch, 1.0 Mb storage ; COMPUTER: Macintosh ; OPERATING SYSTEM: Macintosh ; SOFTWARE: Microsoft Word ; CURRENT APPLICATION DATA: ; APPLICATION NUMBER: US/08/427,640 ; FILING DATE: ; CLASSIFICATION: 435 ; PRIOR APPLICATION DATA: ; APPLICATION NUMBER: 07/689,410 ; FILING DATE: 22 APRIL 1991 ; INFORMATION FOR SEQ ID NO: 5: ; SEQUENCE CHARACTERISTICS: ; LENGTH: 1065 base pairs ; TYPE: nucleic acid ; STRANDEDNESS: double ; TOPOLOGY: linear ; MOLECULE TYPE: DNA US-08-427-640-5</p>			
Qy	1028	AGGCCACATCTGTGTGGGGCGCTGATCCACCCCTGCTGGGTGCTCACTGCTGCCCACT	1087
Db	392	AGCGGTTCCTGTGCGGGGGCATCTCATCAGCTCCTCTGTGGATCTCTCTGCCGCCCACT	451
Qy	1088	GCACCGA-----CATAAAACACAGACATCTAAAGTGTGTCTAGGGGACGAGCACTGA	1141
Db	452	GCITTCAGGAGAGGTTCGCCCCCACCACCTGACGGTGATCTTTGGGCGAACAATACCGGG	511
Qy	1142	AGAAAGAAGATTTTCATGACAGAGCTTTAGGGTGGAAGATATTCAGTACAGCCACT	1201
Db	512	TGTTCCCTCGCGAGGAGGACAGAAAATTTTGAAGTCGAAAATATCATTTGTCCTAAGGAAT	571
Qy	1202	ACAATGAAGAGATGAGATTCGCCACAATCATATTTGCATTGCTCAAGTTAAAGCCAGTGG	1261
Db	572	TCGAT-----GATGACACTTACGACATGACATTTGGCTGCTGCAGCTGAATCGATT	625
Qy	1262	ATGTCTACTGTGCTTAGAATCCAAATACGTGAAGACTGTGTGCTTCCCTGATGGGCTCT	1321

Db 626 CGTCCCGCTGTGCCAGGAGACGCGTGTCCGCACTGTGTGCTTCCCGCGGACC 685
Qy 1322 TTCCCT-----CTGGAGTGTAGTGTCCACATCTCTGCTGGGTGTACAGAAA 1369
Db 686 TGCAGCTGCCGACTGTGAGCTCTCCGGCTACGCGAAGCATGAGGCTTTGT 745
Qy 1370 CAGGAAAGGGTCCCGCAGCTCCCTGATGCCAAAGTCAAGTGTGCAACACTTTGT 1429
Db 746 CTCTTTCTATTCTGGAGCGCTGAGGAGCTCATGTGAGACTGTACCCATCCAGCCGCT 805
Qy 1430 GCAACTCCCGCAACTCTATGACCAACATGATTGATGACAGTATGATCTGTGAGAGAAATC 1489
Db 806 GCACATCACAACTTACTTAACAGAAACAGTACCGACAACTGCTGTGTGTGTGAGACA 865
Qy 1490 TTCAAGAACCTGG-----GCAAGACACCTGCCAGGCTGACTCTGAGGCC 1534
Db 866 CTGGAGCGCGCGCGCCAGGCAACTTGCACGACCTTCCAGGCGATTCGGAGGCC 925
Qy 1535 CCCTGACCTGTGAGAGGAGCGCACCTACTACTGCTATGCGGATGATGAGCTGGGCGCTGG 1594
Db 926 CCCTGTGTGTCTGAACGATGGCGCATGACTTTGGTGGCATCATCAGCTGGGCGCTGG 985
Qy 1595 AGTGTGAGAGAG-----CCAGGGGTGTACACCAAGTTACCAATTCCTGAATTGGA 1648
Db 986 GCTGTGACAGAGAGGATGCCGGGTGTACACCAAGTTACCAACTACCTAGACTGGA 1045
Qy 1649 TCAAGCCACCAT 1661
Db 1046 TTCGTGACACAT 1058

RESULT 11
US-08-427-640-3
; Sequence 3, Application US/08427640
; Patent No. 5658788
; GENERAL INFORMATION:
; APPLICANT: Berg et al.
; TITLE OF INVENTION: Compounds and Methods for Treatment of Thromboembolic Disor
; NUMBER OF SEQUENCES: 28
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Eli Lilly and Company
; STREET: Lilly Corporate Center
; CITY: Indianapolis
; STATE: IN
; COUNTRY: U.S.A.
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.0 Mb storage
; COMPUTER: Macintosh
; OPERATING SYSTEM: Macintosh
; SOFTWARE: Microsoft Word
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/427,640
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/689,410
; FILING DATE: 22 APRIL 1991
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1068 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
US-08-427-640-3

Query Match 4.3%; Score 72.2; DB 1; Length 1068;
Best Local Similarity 50.5%; Pred.No.3.7e-12;
Matches 340; Conservative 0; Mismatches 288; Indels 45; Gaps 5;

Qy 1028 AGGGCCACTTCTGTGTGGGGGCTGATCCACCCCTGCTGGGTGCTCACTGCTGCCCACT 1087

Db 395 AGCGTTCTGTGCGGGGCATACATCAGCTCTCTGCTGGATTTCTCTGCGGCCACT 454
Qy 1088 GCACCGA-----CATAAAAACAGACATCTAAAGTGTGTCTAGGGACACAGGACTGA 1141
Db 455 GCTTCAGAGAGAGGTTTCCGCCCCACACCTGACGCTGATCTTGGGCAACATACCGG 514
Qy 1142 AGAAGAAGAAATTTATGAGCAGAGCTTTAGGGTGAGAGATATTTCAAGTAGAGCCACT 1201
Db 515 TGGTCCCTGCGAGGAGGAGAGAAATTTGAAGTCGAAAAATACATTTGCTCCATAAGAA 574
Qy 1202 ACATGAAAGAGATGAGATTTCCCAACAATGATATTCATTGCTCAAGTTAAAGCCAGTGG 1261
Db 575 TCGAT-----GATGACACTTACGACATGACATTCGGCTGCTGACGTGAAATCGA 628
Qy 1262 ATGCTCACTGCTCTAGAAATCCAAATACGTGAAGACTGTGTGCTGCTGCTGATGGTCT 1321
Db 629 CGTCCCGCTGTGCCAGAGAGAGCGGTGTCGCGCACTGTGTGCTTCCCGCGGAGCC 688
Qy 1322 TTCCCT-----CTGGAGTGTGAGTGCCACATCTCTGCTGCGGTGTTACAGAAA 1369
Db 689 TGCAGCTGCCGAGCTGAGCGAGTGTGAGCTCTCCGGCTACGCAAGCATGAGGCTTGT 748
Qy 1370 CAGAAAAAGGCTCCCGCAGCTCTCTGATCCCAAGTCAAGCTGATGTCACCACTTTGT 1429
Db 749 CTCCTTTCTATTCCGAGCGGCTGAAGGAGGCTCATGTGACACTGTACCCATCCAGCGCT 808
Qy 1430 GCAACTCCCGCAACTCTATGACCAATGATGATGACATGATGATGATGATGATGATGAT 1489
Db 809 GCACATCACAACTTTACTTTAAGAGACAGTCAAGGAGGCTGATGATGATGATGATGAT 868
Qy 1490 TTCAGAAACCTGG-----GCAAGACACCTCCAGGCTGATCTGCGAGGCC 1534
Db 869 CTCGAGCGCGCGCGCCAGGCAAACTTGCACAGCGCTGCCAGGCGATTCGGAGGCC 928
Qy 1535 CCCTGACCTGTGAGAGAGCGGACCTACTACGTCTATGGATGATGATGATGATGATGATG 1594
Db 929 CCCTGGTGTGTCTGAACGATGGCGCATGACTTTGGTGGGCATCATCAGCTGGGCGCT 988
Qy 1595 AGTGTGAGAGAG-----CCAGGGGTGTACACCAAGTTACCAATTCCTGAATTGGA 1648
Db 989 GCTGTGACAGAGAGATGTCGCGGTGTGTACACCAAGTTACCAACTACCTAGACTGGA 1048
Qy 1649 TCAAGCCACCAT 1661
Db 1049 TTCGTGACACAT 1061

RESULT 12
US-09-553-498-9
; Sequence 9, Application US/09553498
; Patent No. 6309861
; GENERAL INFORMATION:
; APPLICANT: Ambrosius, Dorothee
; APPLICANT: Rudolph, Rainer
; APPLICANT: Schaeffner, Joerg
; APPLICANT: Schwarz, Elisabeth
; TITLE OF INVENTION: Process for the production of naturally folded and secreted prot
; FILE REFERENCE: Case 20379
; CURRENT APPLICATION NUMBER: US/09/553,498
; CURRENT FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: EP99107412.1
; PRIOR FILING DATE: 1999-04-26
; NUMBER OF SEQ ID NOS: 10
; SEQ ID NO 9
; LENGTH: 1137
; TYPE: DNA
; ORGANISM: E. coli
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1137)
US-09-553-498-9


```

Query Match      4.3%; Score 72.2; DB 4; Length 1137;
Best Local Similarity 50.5%; Pred. No. 3.8e-12;
Matches 340; Conservative 0; Mismatches 288; Indels 45; Gaps 5;

QY 1028 AGGGCCACTTCTGTGGGGCGCTGATCCACCCCTGCTGGGTGCTCACTGCTGCCACT 1087
DB 461 AGCGGTTCTGTGGGGCGCTATCATCAGCTCCTGTGGATTCTCTGCGGCCACT 520

QY 1088 GCACCGA-----CATAAAAACAGACATCTAAAGGTGGTCTAGGGGACAGGACCTGA 1141
DB 521 GCTTCCAGGAGAGTTTCCGCCCCACACCTGACGCTGATCTTGGGCGAGAACATACCGGG 580

QY 1142 AGAAAGAGAATTTTCATGAGCAGAGCTTTAGGGTGCAGAGATATTCAGTAGGCCACT 1201
DB 581 TGGTCCCTGGGAGGAGGAGAGAAATTTGAAGTCGAAAAATACATTTGTCATAGGAAT 640

QY 1202 ACAATGAAGAGATGAGATTCCCAATATATTCATCTCAAGTTAAAGCCAGTGG 1261
DB 641 TCGAT-----GATGACACTTACGCAATGACATTCGCTGCTCAGCTGAATCGGATT 694

QY 1262 ATGTCACCTGTCTAGAAATCCAAATACGTGAAGACTGTGTGCTGCTGATGGTCT 1321
DB 695 CGTCCCGCTGTGCCAGGAGAGCAGCTGGTCCGCACTGTGTGCTTCCCCCGCGGACC 754

QY 1322 TTCCT-----CTGGAGTGTGCTGCACTCTCTGCTGGGTGTTCAGAAA 1369
DB 755 TGCAGCTCCGAGCTGACGAGTGTGAGCTCTCCGGCTACGGCAGCATGAGCCCTGT 814

QY 1370 CAGGAAAGGGTCCCGCAGCTCTGATGCCAAAGTCAAGCTGATTCGCAACACTTTGT 1429
DB 815 CTCCTTTCTATTTCGGAGCGGTGAAGGAGGCTCATGTGACAGCTGTACCCATCCAGCCGCT 874

QY 1430 GCAACTCCGCGCACTCTATGACCAATGATGACAGTATGATCTGTGCGAGGAATC 1489
DB 875 GCACATCAACAATTTACTTAAAGAACAGTCAACGCAACATGCTGTGTGCTGGAGACA 934

QY 1490 TTCAGAAACCTCG-----GCAAGACACCTGCGAGGCTGACTCTGGAGGCC 1534
DB 935 CTCGGAGCGCGCGCCAGGCAAACTTGCACGACCTGCGAGCGGATTCGGAGGCC 994

QY 1535 CCTGACCTGTGAGAGGAGCGCACTACTAGCTATGGATAGTGTGAGCTGGGCGCTGG 1594
DB 995 CCTGTGTGTCTGAGACGATGGCGCATGACTTTGGTGGGCATCATCAGCTGGGCGCTGG 1054

QY 1595 AGTGTGAGAGAGG-----CCAGGGGTCTACACCAAGTTACCAAAATTCCTGAATTGGA 1648
DB 1055 GCTGTGACAGAGAGTGTCCCGGGTGTGTACACCAAGTTACCAAACTACCTAGACTGGA 1114

QY 1649 TCAAGCCACCAT 1661
DB 1115 TTGCTGACAAAT 1127

```

```

RESULT 13
US-09-618-869-9
; Sequence 9, Application US/09618869
; Patent No. 6455279
; GENERAL INFORMATION:
; APPLICANT: Ambrosius, Dorthée
; APPLICANT: Rudolph, Rainer
; APPLICANT: Schwarzer, Joerg
; APPLICANT: Schwarz, Elisabeth
; TITLE OF INVENTION: PROCESS FOR THE PRODUCTION OF NATURALLY FOLDED AND
; TITLE OF INVENTION: SECRETED PROTEINS BY CO-SECRETION OF MOLECULAR
; TITLE OF INVENTION: CHAPERONES
; FILE REFERENCE: 20381
; CURRENT APPLICATION NUMBER: US/09/618,869
; CURRENT FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: EP99114611.5
; PRIOR FILING DATE: 1999-07-29
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 9

```

```

; LENGTH: 1137
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1137)
; US-09-618-869-9

Query Match      4.3%; Score 72.2; DB 4; Length 1137;
Best Local Similarity 50.5%; Pred. No. 3.8e-12;
Matches 340; Conservative 0; Mismatches 288; Indels 45; Gaps 5;

QY 1028 AGGGCCACTTCTGTGGGGCGCTGATCCACCCCTGCTGGGTGCTCACTGCTGCCACT 1087
DB 461 AGCGGTTCTGTGGGGCGCTATCATCAGCTCCTGTGGATTCTCTGCGGCCACT 520

QY 1088 GCACCGA-----CATAAAAACAGACATCTAAAGGTGGTCTAGGGGACAGGACCTGA 1141
DB 521 GCTTCCAGGAGAGTTTCCGCCCCACACCTGACGCTGATCTTGGGCGAGAACATACCGGG 580

QY 1142 AGAAAGAGAATTTTCATGAGCAGAGCTTTAGGGTGCAGAGATATTCAGTAGGCCACT 1201
DB 581 TGGTCCCTGGGAGGAGGAGAGAAATTTGAAGTCGAAAAATACATTTGTCATAGGAAT 640

QY 1202 ACAATGAAGAGATGAGATTCCCAATATATTCATCTCAAGTTAAAGCCAGTGG 1261
DB 641 TCGAT-----GATGACACTTACGCAATGACATTCGCTGCTCAGCTGAATCGGATT 694

QY 1262 ATGTCACCTGTCTAGAAATCCAAATACGTGAAGACTGTGTGCTGCTGATGGTCT 1321
DB 695 CGTCCCGCTGTGCCAGGAGAGCAGCTGGTCCGCACTGTGTGCTTCCCCCGCGGACC 754

QY 1322 TTCCT-----CTGGAGTGTGCTGCACTCTCTGCTGGGTGTTCAGAAA 1369
DB 755 TGCAGCTCCGAGCTGACGAGTGTGAGCTCTCCGGCTACGGCAGCATGAGCCCTGT 814

QY 1370 CAGGAAAGGGTCCCGCAGCTCTGATGCCAAAGTCAAGCTGATTCGCAACACTTTGT 1429
DB 815 CTCCTTTCTATTTCGGAGCGGTGAAGGAGGCTCATGTGACAGCTGTACCCATCCAGCCGCT 874

QY 1430 GCAACTCCGCGCACTCTATGACCAATGATGACAGTATGATCTGTGCGAGGAATC 1489
DB 875 GCACATCAACAATTTACTTAAAGAACAGTCAACGCAACATGCTGTGTGCTGGAGACA 934

QY 1490 TTCAGAAACCTCG-----GCAAGACACCTGCGAGGCTGACTCTGGAGGCC 1534
DB 935 CTCGGAGCGCGCGCCAGGCAAACTTGCACGACCTGCGAGCGGATTCGGAGGCC 994

QY 1535 CCTGACCTGTGAGAGGAGCGCACTACTAGCTATGGATAGTGTGAGCTGGGCGCTGG 1594
DB 995 CCTGTGTGTCTGAGACGATGGCGCATGACTTTGGTGGGCATCATCAGCTGGGCGCTGG 1054

QY 1595 AGTGTGAGAGAGG-----CCAGGGGTCTACACCAAGTTACCAAAATTCCTGAATTGGA 1648
DB 1055 GCTGTGACAGAGAGTGTCCCGGGTGTGTACACCAAGTTACCAAACTACCTAGACTGGA 1114

QY 1649 TCAAGCCACCAT 1661
DB 1115 TTGCTGACAAAT 1127

```

```

RESULT 14
US-08-811-949-48
; Sequence 48, Application US/08811949
; Patent No. 5840533
; GENERAL INFORMATION:
; APPLICANT: NIWA, MINEO
; APPLICANT: SAITO, YOSHIMASA
; APPLICANT: SASAKI, HITOSHI
; APPLICANT: HAYASHI, MASAKO
; APPLICANT: NOTANI, JOUJI
; APPLICANT: KOBAYASHI, MASAKAZU
; TITLE OF INVENTION: TISSUE PLASMINOGEN ACTIVATOR

```

NUMBER OF SEQUENCES: 67
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MAIER & NEUSTADT,
 ADDRESSEE: P.C.
 STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
 CITY: ARLINGTON
 STATE: VA
 COUNTRY: USA
 ZIP: 22202

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent in Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/811,949
 FILING DATE: 05-MAR-1997
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: OBLON, NORMAN F.
 REGISTRATION NUMBER: 24,618
 REFERENCE/DOCKET NUMBER: 18-966-0
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 703-413-3000
 TELEFAX: 703-413-2220
 INFORMATION FOR SEQ ID NO: 48:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1314 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: circular
 MOLECULE TYPE: DNA (genomic)
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 1..1311

US-08-811-949-48
 Query Match 4.3%; Score 72.2; DB 2; Length 1314;
 Best Local Similarity 50.5%; Pred. No. 4.2e-12;
 Matches 340; Conservative 0; Mismatches 288; Indels 45; Gaps 5;

QY 1028 AGGCCATTCTGTGGTGGGGGCTGATCCACCCCTGCTGGTCTCACTGCTGCCCACT 1087
 DB 638 AGCGTTCTGTGGGGGGGCTACTATCAGCTCTGCTGATCTCTCTGCCCACT 697
 QY 1088 GCACCGA-----CATAAAAACAGACATCTAAAGGTGGTCTAGGGACACGAGCTGA 1141
 DB 698 GTTCCAGGAGAGGTTTCCGCCACCACCTGACGGTGATCTTGGCAGAACATACCGGG 757
 QY 1142 AGAAGAGAGATTTTATGACGACGAGCTTTAGGGTGCAGAGATTTCAAGTACAGCCACT 1201
 DB 758 TGGTCCCTGGGAGGAGGAGAGAGAAATTGAGTGCAGAAATATACATTTCCATAGGAAT 817
 QY 1202 ACAATGAAGAGATGAGATTTCCCAATGATTTGCTCAAGTTAAAGCCAGTGG 1261
 DB 818 TCGAT-----GATGACACTTACGACATGACATGCTGCTGACGCTGAAATCGGATT 871
 QY 1262 ATGGTCACTGTGCTTAGAATCAATAGTGAAGACTGTGCTTGTGCTGATGGTCT 1321
 DB 872 CGTCCCGCTGTGCCAGGAGAGACGGTGGTCCGACATGTGTGCTTCCCGCGCGGACC 931
 QY 1322 TTCCCT-----CTGGGAGTGAAGTGCACATCTCTGGTGGGGTGTACAGAAA 1369
 DB 932 TGCAGTCCGGAGCTGACGAGAGTGTGAGCTCTCCGGCTACGCAAGCATGAGGCTTGT 991
 QY 1370 CAGGAAAGGGTCCCGCCAGCTCTGATGTCGAAAGTCAAGTGAATTCAGTTCGCAACATTGT 1429
 DB 992 CTCCTTTCTATTTCGGAGCGGCTGAAGGAGGCTCATGTGACAGCTGTACCCATCCAGCCGCT 1051
 QY 1430 GGAATCCCGCACTCTATGACCATGATTGATGACAGATGATCTGTGTCAGGAAATC 1489
 DB 1052 GCATATCAACAACTTACTTAAACAGACAGTACCAGACACATGCTGTGTGTCGAGACA 1111

QY 1490 TTCAGAAACCTGG-----GCAAGACACCTGCCAGGTGACTCTTGGAGGCC 1534
 DB 1112 CTGGAGCGGGGGCCCCCAGGCAAACTTGACAGCGCTTCCAGGCGGATTTCGGAGGCC 1171
 QY 1535 CCCTGACCTGTGAGAAAGGACGGACCTTACCTATCTATGGGATAGTGAAGTGGGGCTGG 1594
 DB 1172 CCCTGGTGTCTGACAGATGGCGCATCTTCTTGGGGCATCATCAGCTGGGGCTGG 1231
 QY 1595 AGTGTGAGAAAGG-----CCAGGGGTCTACACCAAGTTACCAAAATTCCTGAATTGA 1648
 DB 1232 GCTGTGACAGAGAGGATGTCCTCCGGTGTGTACAAAGTTACCAACTACCTAGACTGGA 1291
 QY 1649 TCAAAGCCACCAT 1661
 DB 1292 TTCGTGACACAT 1304

RESULT 15
 US-08-883-795A-39
 ; Sequence 39, Application US/08883795A
 ; Patent No. 5985607
 ; GENERAL INFORMATION:
 ; APPLICANT: Delcuve, Genevieve
 ; APPLICANT: Awang, Gregor
 ; TITLE OF INVENTION: Recombinant DNA Molecules and Expression
 ; TITLE OF INVENTION: Vectors for Tissue Plasminogen Activator
 ; NUMBER OF SEQUENCES: 39
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: BERESKIN & PARR
 ; STREET: 40 King Street West
 ; CITY: Toronto
 ; STATE: Ontario
 ; COUNTRY: Canada
 ; ZIP: M5H 3Y2
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent in Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/883,795A
 ; FILING DATE: 27-JUN-1997
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Gravelle, Micheline
 ; REGISTRATION NUMBER: 40,261
 ; REFERENCE/DOCKET NUMBER: 7841-062
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (416) 364-7311
 ; TELEFAX: (416) 361-1398
 ; INFORMATION FOR SEQ ID NO: 39:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1955 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: cDNA
 ; ORIGINAL SOURCE:
 ; ORGANISM: Homo sapiens
 ; US-08-883-795A-39

Query Match 4.3%; Score 72.2; DB 2; Length 1955;
 Best Local Similarity 50.5%; Pred. No. 5.4e-12;
 Matches 340; Conservative 0; Mismatches 288; Indels 45; Gaps 5;
 QY 1028 AGGCCACTTCTGTGGTGGGGCGCTGATCCACCCCTGCTGGTGTCTACTGTGCCCACT 1087
 DB 1103 AGCGTTCTGTGGGGGGCATACTATCAGCTCTCTGATTTCTCTCTGCCGCCACT 1162
 QY 1088 GCACCGA-----CATAAAAACAGACATCTAAGGTGGTGTAGGGACACGAGCTGA 1141
 DB 1163 GCTTCCAGGAGAGGTTTCCGCCCCACCATGCTGCGGTGATCTTGGCAGAACATACCGGG 1222

QY 1142 AGAAGAAGAAATTTTCATGAGAGAGAGCTTTAGGGTGCAGAGATATTTCAAGTACAGCCACT 1201
Db 1223 TGGTCCCTGGCGAGGAGAGAGAAATTTGAAGTCGAAATATACATTTCCATAAGGAAT 1282
QY 1202 ACAAATGAAGAGATGAGATTTCCACAAATGATATTGCAATGCTCAAGTTAAAGCCAGTGG 1261
Db 1283 TCGAT-----GATGACACTTACGACATGACATTTGGCTGCTGCAGCTGAATCGGATT 1336
QY 1262 ATGCTACTGCTGCTAGAAATCCAAATACGTAAGAGACTGTGTGCTTCCCTGATGGTCTCT 1321
Db 1337 CGTCCCGCTGTGCCCAGGAGAGACGCTGTCGCACTGTGTGCTTTCCCCCGCGGACC 1396
QY 1322 TTCCCT-----CTGGGAGTGTGTCGCCACATCTCTGGCTGGGGTGTACAGAAA 1369
Db 1397 TGCAGCTGCCGACTGGACGGAGTGTGAGCTCTCCGGCTACGGCAAGCATGAGGCCCTTGT 1456
QY 1370 CAGGAAAAGGGTCCCGCCAGCTCTGATGTCGAAAGTCAAGCTGATTTGCCAACACTTTGT 1429
Db 1457 CTCCTTTCTATTTCGAGCGGCTGAAGGAGGCTCATGTGACACTGTACCCCATCCAGCCGCT 1516
QY 1430 GCAACTCCCGCCAACTCTATGACCACATGATTGATGACAGTATGATCTGTGCAGGAAATC 1489
Db 1517 GCACATCAACATTTACTTAACAGAACAGTCACCGACACATGCTGTGTGCTGGAGACA 1576
QY 1490 TTCAGAAACCTGG-----GCAAGACACCTGCCAGGGTGAATCTTGGAGGCC 1534
Db 1577 CTGGAGCGGGGGCCCCCAGGCAAACTTGCACGACGCTGCCAGGGGATTCGGGAGGCC 1636
QY 1535 CCCTGACCTGTGAGAGAGACGCGCTACTACTGCTATGGATAGTGAAGTGGGGCCTGG 1594
Db 1637 CCCTGGTGTGTGAACGATGGCGGCATGACTTTGGTGGGCATCATCAGCTGGGGCCTGG 1596
QY 1595 AGTGTGAGAAGAGG-----CCAGGGGTCTACACCCCAAGTTACCAAAATTCCTGAATTGGA 1648
Db 1697 GCTGTGGACAGAGGATGTCCGGGTGTGTACACCAAGTTACCAACTACCTAGACTGGA 1756
QY 1649 TCAAAAGCCACCAT 1661
Db 1757 TTCGTGACACAT 1769

Search completed: May 25, 2004, 10:46:20
Job time : 96.5 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 25, 2004, 09:00:39 ; Search time 528.5 Seconds
(without alignments)
14472.614 Million cell updates/sec

Title: US-09-912-559-2

Perfect score: 1883

Sequence: 1 atgttgccaggatgtctga.....aaagtgaagtggtctctaa 1683

Scoring table:

IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2953838 seqs, 2272363821 residues

Total number of hits satisfying chosen parameters: 5907676

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:**

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
- 2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
- 6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq.*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
- 9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq.*
- 10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq.*
- 11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq.*
- 12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
- 13: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq.*
- 14: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq.*
- 15: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq.*
- 16: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq.*
- 17: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
- 18: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
- 19: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1683	100.0	1683	9	US-09-912-559-2
2	1683	100.0	1683	16	US-10-391-215-4
3	1681.4	99.9	1683	16	US-10-391-215-2
4	1681.4	99.9	1683	16	US-10-391-215-3
5	1679.8	99.8	1683	9	US-09-912-559-1
6	1679.8	99.8	1683	16	US-10-391-215-1
7	1679.8	99.8	3008	9	US-09-880-107-1668
8	368.4	21.9	428	10	US-09-918-995-8242
9	262.4	15.6	264	13	US-10-425-000-17
10	120.2	7.1	451	9	US-09-864-761-11164
11	117	7.0	117	9	US-09-864-761-27791
12	114.4	6.8	2036	9	US-09-954-456-552
13	114.4	6.8	2036	9	US-09-880-107-1612
14	93.2	5.5	1302	13	US-10-087-192-590

Sequence 10, Appl	15	93.2	5.5	2299	13	US-10-665-216-10
Sequence 174961,	c 16	74.4	4.4	3186778	13	US-10-027-632-174961
Sequence 174961,	c 17	74.4	4.4	3186778	16	US-10-027-632-174961
Sequence 33, Appl	18	72.8	4.3	614	9	US-09-969-271-6
Sequence 6, Appl	19	72.2	4.3	1689	9	US-09-969-271-6
Sequence 25, Appl	20	72.2	4.3	1726	13	US-10-411-037-25
Sequence 25, Appl	21	72.2	4.3	1726	13	US-10-411-026-25
Sequence 25, Appl	22	72.2	4.3	1726	17	US-10-410-362-25
Sequence 541, App	23	72.2	4.3	1726	17	US-10-411-049-25
Sequence 541, App	24	72.2	4.3	2461	13	US-10-342-887-541
Sequence 541, App	25	72.2	4.3	2461	13	US-10-172-118-541
Sequence 7, Appl	26	72.2	4.3	2509	13	US-10-193-656-7
Sequence 5, Appl	27	72.2	4.3	2519	9	US-09-969-271-5
Sequence 540, App	28	72.2	4.3	2519	13	US-10-342-887-540
Sequence 540, App	29	72.2	4.3	2519	13	US-10-172-118-540
Sequence 3, Appl	30	72.2	4.3	2544	15	US-10-443-701-3
Sequence 144, Appl	31	72.2	4.3	2641	9	US-09-974-298-144
Sequence 19992, A	32	72.2	4.3	2859	10	US-09-814-353-19992
Sequence 4, Appl	33	70.6	4.2	1065	10	US-09-987-455-4
Sequence 7, Appl	34	70.6	4.2	1065	10	US-09-987-455-7
Sequence 2, Appl	35	70.6	4.2	1128	10	US-09-987-455-2
Sequence 5, Appl	36	70.6	4.2	1128	10	US-09-987-455-5
Sequence 57, Appl	37	70.6	4.2	1996	16	US-10-401-077-2
Sequence 56, Appl	38	69.2	4.1	7596	15	US-10-004-113-57
Sequence 2167, Ap	39	69.2	4.1	8064	15	US-10-004-113-56
Sequence 132, App	40	68.6	4.1	8221	12	US-10-152-319A-2167
Sequence 14421, A	41	67.8	4.0	505	10	US-09-814-353-14421
Sequence 34, Appl	42	67.8	4.0	2160	16	US-10-159-563-132
Sequence 163, App	43	67.4	4.0	2328	15	US-10-247-671-34
Sequence 16, Appl	44	67.4	4.0	2341	13	US-10-116-802-163
	c 45	66.4	3.9	1078	10	US-09-898-837A-16

ALIGNMENTS

RESULT 1

US-09-912-559-2
; Sequence 2, Application US/09912559
; Patent No. US20020142316A1
; GENERAL INFORMATION:
; APPLICANT: ROEMISCH, JUERGEN
; APPLICANT: STOEHR, HANS-ARNOLD
; APPLICANT: FEUSSNER, ANNETTE
; APPLICANT: LANG, WIEGAND
; APPLICANT: WEIMER, THOMAS
; APPLICANT: BECKER, MARGRET
; APPLICANT: NERLICH, CLAUDIA
; APPLICANT: MUTH-NAUMANN, GUDRUN
; TITLE OF INVENTION: MUTANTS OF THE FACTOR VII-ACTIVATING PROTEASE AND
; TITLE OF INVENTION: DETECTION METHODS USING SPECIFIC ANTIBODIES
; FILE REFERENCE: 06478.1457
; CURRENT APPLICATION NUMBER: US/09/912,559
; CURRENT FILING DATE: 2001-07-26
; PRIOR APPLICATION NUMBER: DE 100 36 641.4
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: DE 100 50 040.4
; PRIOR FILING DATE: 2000-10-10
; PRIOR APPLICATION NUMBER: DE 100 52 319.6
; PRIOR FILING DATE: 2000-10-21
; PRIOR APPLICATION NUMBER: DE 101 18 706.8
; PRIOR FILING DATE: 2001-04-12
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 2
; LENGTH: 1683
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-912-559-2

Query Match 100.0%; Score 1683; DB 9; Length 1683;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1683; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGTTGCCAGGATGCTGATCTCCATGTTCTCTGTTAATGGCTCTGTTGGGAAAGACA 60
Db 1 ATGTTGCCAGGATGCTGATCTCCATGTTCTCTGTTAATGGCTCTGTTGGGAAAGACA 60
QY 61 GCCTGTGGGTTCTCCCTGATGCTCTTATTGGAAGCTGGACCCAGACTGGACCCCTGAC 120
Db 61 GCCTGTGGGTTCTCCCTGATGCTCTTATTGGAAGCTGGACCCAGACTGGACCCCTGAC 120
QY 121 CAGTATGATTACAGCTACGAGGATTAATATCAGGAAGAGAACACCAAGTAGCACATTACC 180
Db 121 CAGTATGATTACAGCTACGAGGATTAATATCAGGAAGAGAACACCAAGTAGCACATTACC 180
QY 181 CATGCTGAGAACTCTGACTGCTACTACATGAGGACCAAGCTGATCCATGCCAGCCAC 240
Db 181 CATGCTGAGAACTCTGACTGCTACTACATGAGGACCAAGCTGATCCATGCCAGCCAC 240
QY 241 CCCTGTGAACACGCTGGGGAGCTGCTGCTCCATCGGAGCACCTTACATGCACTGCCCTG 300
Db 241 CCCTGTGAACACGCTGGGGAGCTGCTGCTCCATCGGAGCACCTTACATGCACTGCCCTG 300
QY 301 GCTCCTTCTCTGGGATAGTGTCAAGAGTGCMAATACGTGCAAGGCAACCCCTGCCAG 360
Db 301 GCTCCTTCTCTGGGATAGTGTCAAGAGTGCMAATACGTGCAAGGCAACCCCTGCCAG 360
QY 361 GGCGGGGCAATGCTCTATTACCCAGAGTCCCTTACTACCGCTGTGTCTGTAAACAC 420
Db 361 GGCGGGGCAATGCTCTATTACCCAGAGTCCCTTACTACCGCTGTGTCTGTAAACAC 420
QY 421 CTTTACACAGTCCAGCTGCTCCAGTGGTTCCTGTATGAGGCAACCCCTGCCAG 480
Db 421 CTTTACACAGTCCAGCTGCTCCAGTGGTTCCTGTATGAGGCAACCCCTGCCAG 480
QY 481 AATGGGCTACCTGCTCCCGGATAGCGGAGATCCAAAGTTCACCTGTGCTGCTCCGAC 540
Db 481 AATGGGCTACCTGCTCCCGGATAGCGGAGATCCAAAGTTCACCTGTGCTGCTCCGAC 540
QY 541 CAGTTCAAGGGAAATCTGTGAATAGGTTCTGATGACTGCTATGTGGCGATGGCTAC 600
Db 541 CAGTTCAAGGGAAATCTGTGAATAGGTTCTGATGACTGCTATGTGGCGATGGCTAC 600
QY 601 TCTTACCGAGGAAATGAATAGACAGTCAACAGCATGCTGCTTACTGGAATCTCC 660
Db 601 TCTTACCGAGGAAATGAATAGACAGTCAACAGCATGCTGCTTACTGGAATCTCC 660
QY 661 CACTCTCTTGCAGGAATTAACAATGTTATGAGGATGCTGAAACCCATGGAT 720
Db 661 CACTCTCTTGCAGGAATTAACAATGTTATGAGGATGCTGAAACCCATGGAT 720
QY 721 GGGGAACACAAATTTCTCAGAAACCCAGATGCGGAGCAAGCCCTGGTCTTATAA 780
Db 721 GGGGAACACAAATTTCTCAGAAACCCAGATGCGGAGCAAGCCCTGGTCTTATAA 780
QY 781 GTTACCAATGACAGGTAAGTGAATATCTGTGATGCTCAGCCTGCTCAGCCAGGAC 840
Db 781 GTTACCAATGACAGGTAAGTGAATATCTGTGATGCTCAGCCTGCTCAGCCAGGAC 840
QY 841 GTTSCCTTACCCAGAGGAAGCCCACTGAGCCATCAACCAAGCTCCGGGTTGACTCC 900
Db 841 GTTSCCTTACCCAGAGGAAGCCCACTGAGCCATCAACCAAGCTCCGGGTTGACTCC 900
QY 901 TGTGGAAGACTGAGATAGCAGAGGAGGAATCAAGAGATCTATGAGGCTTTAAGAGC 960
Db 901 TGTGGAAGACTGAGATAGCAGAGGAGGAATCAAGAGATCTATGAGGCTTTAAGAGC 960
QY 961 ACAGGGGCAAGCAACCCATGGCAGGCTCCCTCAGTCTCCTGCTGACCATCTCC 1020
Db 961 ACAGGGGCAAGCAACCCATGGCAGGCTCCCTCAGTCTCCTGCTGACCATCTCC 1020
QY 1021 ATGCCCCAGGGCCACTTCTGTGTGGGGCGCTGATCCACCCCTGCTGGGTGCTCACTGCT 1080
Db 1021 ATGCCCCAGGGCCACTTCTGTGTGGGGCGCTGATCCACCCCTGCTGGGTGCTCACTGCT 1080

QY 1081 GCCCACTGCAACCCAGACATAAAAAACAGACATCTAAAGTGTGCTAGGGGACCAAGCCTG 1140
Db 1081 GCCCACTGCAACCCAGACATAAAAAACAGACATCTAAAGTGTGCTAGGGGACCAAGCCTG 1140
QY 1141 AAGAAAGAGAAATTTTCATGAGCAGAGCTTTAGGGTGCAGAGATATTCAAGTACAGCCAC 1200
Db 1141 AAGAAAGAGAAATTTTCATGAGCAGAGCTTTAGGGTGCAGAGATATTCAAGTACAGCCAC 1200
QY 1201 TACAATGAAGAGATGAGATTCCCAATGATTTGCATTTGCTCAAGTTAAAGTCCAGTG 1260
Db 1201 TACAATGAAGAGATGAGATTCCCAATGATTTGCATTTGCTCAAGTTAAAGTCCAGTG 1260
QY 1261 GATGTCACCTGCTCTAGAAATCCAAATACGTGAAGACTGTGCTTGTGCTGATGGTCC 1320
Db 1261 GATGTCACCTGCTCTAGAAATCCAAATACGTGAAGACTGTGCTTGTGCTGATGGTCC 1320
QY 1321 TTTCCCTCTGGAGTGTGAGTGCACATCTCTGGTGGGGTTACAGAAAACAGAAAAGGG 1380
Db 1321 TTTCCCTCTGGAGTGTGAGTGCACATCTCTGGTGGGGTTACAGAAAACAGAAAAGGG 1380
QY 1381 TCCCGCAGCTCTGGATGCCAAGTCAAGTATTCGCCAACCTTTGTGCAACTCCCGC 1440
Db 1381 TCCCGCAGCTCTGGATGCCAAGTCAAGTATTCGCCAACCTTTGTGCAACTCCCGC 1440
QY 1441 CAACTCTATGACACATGATGATGACAGTATGATCTGTGAGGAAATCTTCAGAAACCT 1500
Db 1441 CAACTCTATGACACATGATGATGACAGTATGATCTGTGAGGAAATCTTCAGAAACCT 1500
QY 1501 GGGCAAGACACTCCAGGCTGACTCTGGAGGCCCTGCTGAGAGGAGCGGACCC 1560
Db 1501 GGGCAAGACACTCCAGGCTGACTCTGGAGGCCCTGCTGAGAGGAGCGGACCC 1560
QY 1561 TACTACGCTCTATGGAGTAGTGAGCTGGGGCTGAGAGTGTGAGAGAGCGGAGCTAC 1620
Db 1561 TACTACGCTCTATGGAGTAGTGAGCTGGGGCTGAGAGTGTGAGAGAGCGGAGCTAC 1620
QY 1621 ACCCAAGTTACAAATTCCTGAATTCGATCAAGCCACCATCAAAAGTGAAGTGGCTTC 1680
Db 1621 ACCCAAGTTACAAATTCCTGAATTCGATCAAGCCACCATCAAAAGTGAAGTGGCTTC 1680
QY 1681 TAA 1683
Db 1681 TAA 1683
RESULT 2
US-10-391-215-4
; Sequence 4, Application US/10391215
; Publication No. US20040009543A1
; GENERAL INFORMATION:
; APPLICANT: KIEHL, STEFAN
; APPLICANT: WILLEIT, JOHANN
; APPLICANT: WIEDERMANN, CHRISTIAN JOSEF
; APPLICANT: ROEMISCH, JUERGEN
; APPLICANT: WEIMER, THOMAS
; APPLICANT: FEUSSNER, ANNETTE
; APPLICANT: STOEHR, HANS-ARNOLD
; APPLICANT: DOERSAM, VOLKER
; TITLE OF INVENTION: MARBURG I MUTANT OF FACTOR VII ACTIVATING PROTEASE
; FILE REFERENCE: 06478.1457-01
; CURRENT APPLICATION NUMBER: US/10/391.215
; CURRENT FILING DATE: 2003-03-19
; PRIOR APPLICATION NUMBER: 09/912,559
; PRIOR FILING DATE: 2001-07-26
; PRIOR APPLICATION NUMBER: DE 100 36 641.4
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: DE 100 50 040.4
; PRIOR FILING DATE: 2000-10-10
; PRIOR APPLICATION NUMBER: DE 100 52 319.6
; PRIOR FILING DATE: 2000-10-21
; PRIOR APPLICATION NUMBER: DE 101 18 706.8
; PRIOR FILING DATE: 2001-04-12

This Page Blank (uspto)

; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 1683
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-391-215-4

Query Match 100.0%; Score 1683; DB 16; Length 1683;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1683; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGTTTCCAGAGTCTGATCTCCATGTTCTGCTGTTAATGCTCTGTTGGGAAGACA 60
Db 1 ATGTTTCCAGAGTCTGATCTCCATGTTCTGCTGTTAATGCTCTGTTGGGAAGACA 60
Qy 61 GCCTGTGGGTCTCCCTGATGTTCTTTATGGAAGCCTGGACCCAGACTGGACCCCTGAC 120
Db 61 GCCTGTGGGTCTCCCTGATGTTCTTTATGGAAGCCTGGACCCAGACTGGACCCCTGAC 120
Qy 121 CAGTATGATTACAGTACAGGAGTATTAATCAGGAAGACACACAGTACACACTTACC 180
Db 121 CAGTATGATTACAGTACAGGAGTATTAATCAGGAAGACACACAGTACACACTTACC 180
Qy 181 CATGCTGAGATCTGCTGATCTACACTGAGGACCAAGCTGATCCATGCCAGCCCAAC 240
Db 181 CATGCTGAGATCTGCTGATCTACACTGAGGACCAAGCTGATCCATGCCAGCCCAAC 240
Qy 241 CCCTGTGAACACGGTGGGACTGCTGCTCATGGGAGACCTTTCATGAGCTGCTG 300
Db 241 CCCTGTGAACACGGTGGGACTGCTGCTCATGGGAGACCTTTCATGAGCTGCTG 300
Qy 301 GCTCTTCTCTGGGAATGATGTCAGAAAGTCAAAATACGTCAAGGACCAACCCATGT 360
Db 301 GCTCTTCTCTGGGAATGATGTCAGAAAGTCAAAATACGTCAAGGACCAACCCATGT 360
Qy 361 GGCGGGCCCAATGCTCATACAGAGTCTCCCTACTACCGCTGCTGTGTTAAACAC 420
Db 361 GGCGGGCCCAATGCTCATACAGAGTCTCCCTACTACCGCTGCTGTGTTAAACAC 420
Qy 421 CCTTACAGGTCAGCTGCTCCAGTGTCTGATGATGATGATGATGATGATGATGATG 480
Db 421 CCTTACAGGTCAGCTGCTCCAGTGTCTGATGATGATGATGATGATGATGATGATG 480
Qy 481 AATGGGGTCACTGCTCCCGGATTAAGCGGAGATCAAGTTCACTGCTGCTGCTCCGAC 540
Db 481 AATGGGGTCACTGCTCCCGGATTAAGCGGAGATCAAGTTCACTGCTGCTGCTCCGAC 540
Qy 541 CAGTTCAAGGGAAATCTGGAATAGTCTGATGATGATGATGATGATGATGATGATG 600
Db 541 CAGTTCAAGGGAAATCTGGAATAGTCTGATGATGATGATGATGATGATGATGATG 600
Qy 601 TCTTACCGAGGAAATGAATAGGACAGTCAACAGCATGCTGCTTTACTGGAATCC 660
Db 601 TCTTACCGAGGAAATGAATAGGACAGTCAACAGCATGCTGCTTTACTGGAATCC 660
Qy 661 CACTCTCTTTCAGGAGAAATTAACATGTTTATGGAGTCTGTAACCATGGGATT 720
Db 661 CACTCTCTTTCAGGAGAAATTAACATGTTTATGGAGTCTGTAACCATGGGATT 720
Qy 721 GGGGAACACAATTTCTGAGAAACCCAGATGGGACGAAAGCCCTGCTGTTTATAA 780
Db 721 GGGGAACACAATTTCTGAGAAACCCAGATGGGACGAAAGCCCTGCTGTTTATAA 780
Qy 781 GTTACCAATGACAAGGTGAAATGGGAATGATGATGATGATGATGATGATGATGATG 840
Db 781 GTTACCAATGACAAGGTGAAATGGGAATGATGATGATGATGATGATGATGATGATG 840
Qy 841 GTTGCTTACCCAGAGAAAGCCCATGAGCCATCAACAGCTTCCGGGGTTGACTCC 900
Db 841 GTTGCTTACCCAGAGAAAGCCCATGAGCCATCAACAGCTTCCGGGGTTGACTCC 900
Qy 901 TGTGGAAAGACTGAGATAGCAGAGAGAAATCAAGAGAAATCTATGGAGGCTTTAAGAC 960

Db 901 TGTGAAAGACTGAGATAGCAGAGAGAAATCAAGAGAAATCTATGAGGCTTTAAGAC 960
Qy 961 ACGCGGGCAAGCAACCATGCGCGCTCCCTCCAGTCTCGCTCCCTGACCATCTCC 1020
Db 961 ACGCGGGCAAGCAACCATGCGCGCTCCCTCCAGTCTCGCTCCCTGACCATCTCC 1020
Qy 1021 ATGCCCCAGGGCCACTTCTGTTGGGGCGCTGATCCACCCCTGCTGGGTCTCACTGCT 1080
Db 1021 ATGCCCCAGGGCCACTTCTGTTGGGGCGCTGATCCACCCCTGCTGGGTCTCACTGCT 1080
Qy 1081 GCCCACTCCACCGAATAAAAAACAGACATCTAAAGTGTGCTAGGGGACAGGACCTG 1140
Db 1081 GCCCACTCCACCGAATAAAAAACAGACATCTAAAGTGTGCTAGGGGACAGGACCTG 1140
Qy 1141 AAGAAGAGAAATTTCTAGGAGAGCTTTAAGGTGCAAGAGATATCAAGTACAGCCAC 1200
Db 1141 AAGAAGAGAAATTTCTAGGAGAGCTTTAAGGTGCAAGAGATATCAAGTACAGCCAC 1200
Qy 1201 TACAATGAAGAGATGAGATTTCCCAATGATATGATGCTCAAGTTAAAGCAGT 1260
Db 1201 TACAATGAAGAGATGAGATTTCCCAATGATATGATGCTCAAGTTAAAGCAGT 1260
Qy 1261 GATGTCATGCTGCTAGTATCCAAATACGTTGAAGACTGTTGCTTCCCTGATGGTCC 1320
Db 1261 GATGTCATGCTGCTAGTATCCAAATACGTTGAAGACTGTTGCTTCCCTGATGGTCC 1320
Qy 1321 TTTTCCCTCTGGGAGTGAAGTCCCAATCTCTGGCTGGGTGTTACAGAAACAGGAAAGG 1380
Db 1321 TTTTCCCTCTGGGAGTGAAGTCCCAATCTCTGGCTGGGTGTTACAGAAACAGGAAAGG 1380
Qy 1381 TCCCGCAGCTCTGATGAGTCCAAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGT 1440
Db 1381 TCCCGCAGCTCTGATGAGTCCAAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGT 1440
Qy 1441 CAACTCTATGACACATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1500
Db 1441 CAACTCTATGACACATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1500
Qy 1501 GGGCAAGACCTGCGAGGTCATCTGGAGGCCCCCTGACCTGTGAGAGGACGCGCAC 1560
Db 1501 GGGCAAGACCTGCGAGGTCATCTGGAGGCCCCCTGACCTGTGAGAGGACGCGCAC 1560
Qy 1561 TACTAGTCTATGGGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1620
Db 1561 TACTAGTCTATGGGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1620
Qy 1621 ACCCAAGTTACCAATTTCTGATGATGATGATGATGATGATGATGATGATGATGATG 1680
Db 1621 ACCCAAGTTACCAATTTCTGATGATGATGATGATGATGATGATGATGATGATGATG 1680
Qy 1681 TAA 1683
Db 1681 TAA 1683

RESULT 3

US-10-391-215-2
; Sequence 2, Application US/10391215
; Publication NO. US20040009543A1
; GENERAL INFORMATION:
; APPLICANT: KIEHL, STEFAN
; APPLICANT: WILLEIT, JOHANN
; APPLICANT: WIEDERMANN, CHRISTIAN JOSEF
; APPLICANT: ROEMISCH, JUERGEN
; APPLICANT: WEIMER, THOMAS
; APPLICANT: FEUSSNER, ANNETTE
; APPLICANT: STOEHR, HANS-ARNOLD
; APPLICANT: DOBERSAM, VOLKER
; TITLE OF INVENTION: MARBURG I MUTANT OF FACTOR VII ACTIVATING PROTEASE
; TITLE OF INVENTION: (FSAP) AS RISK FACTOR FOR ARTERIAL THROMBOSIS
; FILE REFERENCE: 06478.1457-01
; CURRENT APPLICATION NUMBER: US/10/391.215

CURRENT FILING DATE: 2003-03-19
; PRIOR APPLICATION NUMBER: 09/912,559
; PRIOR FILING DATE: 2001-07-26
; PRIOR APPLICATION NUMBER: DE 100 36 641.4
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: DE 100 50 040.4
; PRIOR FILING DATE: 2000-10-10
; PRIOR APPLICATION NUMBER: DE 100 52 319.6
; PRIOR FILING DATE: 2000-10-21
; PRIOR APPLICATION NUMBER: DE 101 18 706.8
; PRIOR FILING DATE: 2001-04-12
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 1683
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-391-215-2

Query Match 99.9%; Score 1681.4; DB 16; Length 1683;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1682; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY	1	ATGTTTCCAGGATGCTGATCTCCATCTGTTCTGCTGTTAATGCTCTGTTGGGAAAGACA	60
DB	1	ATGTTTCCAGGATGCTGATCTCCATCTGTTCTGTTAATGCTCTGTTGGGAAAGACA	60
QY	61	GCTGTGGGTTCTCCCTGATGTTTATTGGAAGCTGGACCCAGACTGGACCCCTGAC	120
DB	61	GCTGTGGGTTCTCCCTGATGTTTATTGGAAGCTGGACCCAGACTGGACCCCTGAC	120
QY	121	CAGTATGATTACAGTACAGGAGGATTATATCAGGAGAGAACACCAAGTACACATTACC	180
DB	121	CAGTATGATTACAGTACAGGAGGATTATATCAGGAGAGAACACCAAGTACACATTACC	180
QY	181	CATGCTCAGGAATCTGACTGTTACTACACTGAGGACCAAGCTGATCCATGCCAGCCCAAC	240
DB	181	CATGCTCAGGAATCTGACTGTTACTACACTGAGGACCAAGCTGATCCATGCCAGCCCAAC	240
QY	241	CCCTGTGAAACAGGTGGGAGTCTGCTGTCATGAGGAGCCTTACATGCACTGCCCTG	300
DB	241	CCCTGTGAAACAGGTGGGAGTCTGCTGTCATGAGGAGCCTTACATGCACTGCCCTG	300
QY	301	GCTCCTTTCTCTGGGAATAGTGTGAGAAAGTGCAGAAATACGTGCAAGGACAAACCCATGT	360
DB	301	GCTCCTTTCTCTGGGAATAGTGTGAGAAAGTGCAGAAATACGTGCAAGGACAAACCCATGT	360
QY	361	GGCGGGGCCAATGCTCATTACCCAGAGTCTCCCTACTACCGCTGTGTGTAAACAC	420
DB	361	GGCGGGGCCAATGCTCATTACCCAGAGTCTCCCTACTACCGCTGTGTGTAAACAC	420
QY	421	CCTTACAGAGTCCAGCTGCTCCCAAGTGTCTGTATGAGGCGCAAAACCCCTGCCAG	480
DB	421	CCTTACAGAGTCCAGCTGCTCCCAAGTGTCTGTATGAGGCGCAAAACCCCTGCCAG	480
QY	481	AATGGGGTACTGCTCCCGGCATAAGCGGAGATCCAGTTACCTGTGCTGCCGAC	540
DB	481	AATGGGGTACTGCTCCCGGCATAAGCGGAGATCCAGTTACCTGTGCTGCCGAC	540
QY	541	CAGTTCAAGGGGAAATCTGTGAATAGTTTCTGATGACTGCTATGTTGGGATGGCTAC	600
DB	541	CAGTTCAAGGGGAAATCTGTGAATAGTTTCTGATGACTGCTATGTTGGGATGGCTAC	600
QY	601	TCTTACCGAGGAAATGAATAGGACACTCAACGAGATGCGTGCCTTTACTGGAATCC	660
DB	601	TCTTACCGAGGAAATGAATAGGACACTCAACGAGATGCGTGCCTTTACTGGAATCC	660
QY	661	CACCTCTCTTCAGGAGGATTAACATGTTTATGGAGGATGCTGAAACCCATGGGATT	720
DB	661	CACCTCTCTTCAGGAGGATTAACATGTTTATGGAGGATGCTGAAACCCATGGGATT	720
QY	721	GGGGAACCAATTTCTGCAGAAACCCAGATCGGACGAAAGCCCTGCTGTTTATTAA	780
DB	721	GGGGAACCAATTTCTGCAGAAACCCAGATCGGACGAAAGCCCTGCTGTTTATTAA	780

DB	721	GGGGAACCAATTTCTGCAGAAACCCAGATCGGACGAAAGCCCTGCTGTTTATTAA	780
QY	781	GTTACCAATGACAAAGTGAATGGAATACTGTGATGCTTACGCTGCTCAGCCAGGAC	840
DB	781	GTTACCAATGACAAAGTGAATGGAATACTGTGATGCTTACGCTGCTCAGCCAGGAC	840
QY	841	GTTCCCTTACCGAGGAAAGCCCACTGAGCCTCAACCAAGCTTCGGGGTTTGACTCC	900
DB	841	GTTCCCTTACCGAGGAAAGCCCACTGAGCCTCAACCAAGCTTCGGGGTTTGACTCC	900
QY	901	TGTGAAAGACTGAGATAGCAGAGAGAGATCAAGAGATATCTATGGAGCTTTAAGAGC	960
DB	901	TGTGAAAGACTGAGATAGCAGAGAGAGATCAAGAGATATCTATGGAGCTTTAAGAGC	960
QY	961	ACGGGGGCAAGCAACCATGCGAGGCGTCCCTCAGTCCCTCGTCTGACCATCTCC	1020
DB	961	ACGGGGGCAAGCAACCATGCGAGGCGTCCCTCAGTCCCTCGTCTGACCATCTCC	1020
QY	1021	ATGCCCCAGGCGCACTTCTGTGGTGGGCGCTGATCCACCCCTGCTGGTGGTCTC	1080
DB	1021	ATGCCCCAGGCGCACTTCTGTGGTGGGCGCTGATCCACCCCTGCTGGTGGTCTC	1080
QY	1081	GCCCACTGCACCGACATAAATAACAGACATCTAAGGTGCTAGGGGACCGAGCCTG	1140
DB	1081	GCCCACTGCACCGACATAAATAACAGACATCTAAGGTGCTAGGGGACCGAGCCTG	1140
QY	1141	AAGAAAGAGAAATTTCTGATGAGCAGCTTTAGGGTGCAGAGATATTTCAAGTACAGCCAC	1200
DB	1141	AAGAAAGAGAAATTTCTGATGAGCAGCTTTAGGGTGCAGAGATATTTCAAGTACAGCCAC	1200
QY	1201	TACAAATGAAAGAGATGAGATTCCCAACAATGATATTCATTTGCTCAAGATTAAGCCAGTG	1260
DB	1201	TACAAATGAAAGAGATGAGATTCCCAACAATGATATTCATTTGCTCAAGATTAAGCCAGTG	1260
QY	1261	GATGGTCACTGCTCTAGATCCAAATACGTGAAGACTGTGTGCTTGCCTGATGGGTCC	1320
DB	1261	GATGGTCACTGCTCTAGATCCAAATACGTGAAGACTGTGTGCTTGCCTGATGGGTCC	1320
QY	1321	TTTCCCTCTGGGAGTGCACATCTCTGGTGGGGTGTTCAGAAAACAGGAAAAGGG	1380
DB	1321	TTTCCCTCTGGGAGTGCACATCTCTGGTGGGGTGTTCAGAAAACAGGAAAAGGG	1380
QY	1381	TCCCGCAGCTCTGGATGCCAAGTCAAGCTGATTCGCAACACTTTGTGCAACTCCGCG	1440
DB	1381	TCCCGCAGCTCTGGATGCCAAGTCAAGCTGATTCGCAACACTTTGTGCAACTCCGCG	1440
QY	1441	CAACTCTATGACCACATGATTGATGACAGATATGATCTGTGAGGAAATCTTTCAGAACT	1500
DB	1441	CAACTCTATGACCACATGATTGATGACAGATATGATCTGTGAGGAAATCTTTCAGAACT	1500
QY	1501	GGGCAAGACACTGCGAGGCTGACTCTGGAGGCGCCCTGACCTGTGAGAGAGCGGACCC	1560
DB	1501	GGGCAAGACACTGCGAGGCTGACTCTGGAGGCGCCCTGACCTGTGAGAGAGCGGACCC	1560
QY	1561	TACTACGCTCTATGGGATAGTGGGCGCTGGAGTGTGAGAGAGCGGCGGTCTTAC	1620
DB	1561	TACTACGCTCTATGGGATAGTGGGCGCTGGAGTGTGAGAGAGCGGCGGTCTTAC	1620
QY	1621	ACCCAGTTTACAAATCTCTGAATGGAATCAAGCCACCATCAAGAGTGAAGTGGCTTC	1680
DB	1621	ACCCAGTTTACAAATCTCTGAATGGAATCAAGCCACCATCAAGAGTGAAGTGGCTTC	1680
QY	1681	TAA 1683	
DB	1681	TAA 1683	

RESULT 4
US-10-391-215-3
; Sequence 3, Application US/10391215
; Publication No. US20040009543A1
; GENERAL INFORMATION:
; APPLICANT: KIECHUL, STEFAN

APPLICANT: WILLEIT, JOHANN
APPLICANT: WIEDERMANN, CHRISTIAN JOSEF
APPLICANT: ROEMISCH, JUERGEN
APPLICANT: WEIMER, THOMAS
APPLICANT: FEUSSNER, ANNETTE
APPLICANT: STOEHR, HANS-ARNOLD
APPLICANT: DOERSAM, VOLKER
TITLE OF INVENTION: MARBURG I MUTANT OF FACTOR VII ACTIVATING PROTEASE
FILE REFERENCE: (FSP) AS RISK FACTOR FOR ARTERIAL THROMBOSIS
CURRENT APPLICATION NUMBER: US/10/391.215
CURRENT FILING DATE: 2003-03-19
PRIORITY APPLICATION NUMBER: 09/912,559
PRIORITY FILING DATE: 2001-07-26
PRIORITY FILING DATE: 2000-07-26
PRIORITY FILING DATE: 2000-07-26
PRIORITY FILING DATE: 2000-10-10
PRIORITY FILING DATE: 2000-10-21
PRIORITY FILING DATE: 2001-04-12
NUMBER OF SEQ ID NOS: 8
SOFTWARE: Patent In Ver. 2.1
SEQ ID NO 3
TYPE: DNA
ORGANISM: Homo sapiens
US-10-391-215-3

Query Match 99.9%; Score 1681.4; DB 16; Length 1683;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1682; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ATGTTTCCAGAGTCTGATCTCATTCTGCTGTTTATGGTCTCTGGTGGGAAGACA 60
Db 1 ATGTTTCCAGAGTCTGATCTCATTCTGCTGTTTATGGTCTCTGGTGGGAAGACA 60

Qy 61 GCCTGTGGGTTCTCCCTGATGCTTTATTGAAAGCCCTGGACCCGACCTGGAC 120
Db 61 GCCTGTGGGTTCTCCCTGATGCTTTATTGAAAGCCCTGGACCCGACCTGGAC 120

Qy 121 CAGTATGATTACAGTCTCAGAGTATTAATCAGGAAGACACACAGTAGCACACTTACC 180
Db 121 CAGTATGATTACAGTCTCAGAGTATTAATCAGGAAGACACACAGTAGCACACTTACC 180

Qy 181 CATGCTGAGATCCCTGACTGTGTTACTACTGAGGACCAAGCTGATCCATGCGCCCAAC 240
Db 181 CATGCTGAGATCCCTGACTGTGTTACTACTGAGGACCAAGCTGATCCATGCGCCCAAC 240

Qy 241 CCCTGTGAACAGGTGGGACTGCTCTGTCATGGGAGACCTTACATGAGCTGGCTG 300
Db 241 CCCTGTGAACAGGTGGGACTGCTCTGTCATGGGAGACCTTACATGAGCTGGCTG 300

Qy 301 GCTCCTTTCTCTGGGAATAAGTGTGAGAAAGTGCAGAAATAGCTGCAAGGACAAACCCATGT 360
Db 301 GCTCCTTTCTCTGGGAATAAGTGTGAGAAAGTGCAGAAATAGCTGCAAGGACAAACCCATGT 360

Qy 361 GCGCGGGCCAAATGCTCATTAACAGAGTCTCCTACTACCGTGTGCTGTGTAACAC 420
Db 361 GCGCGGGCCAAATGCTCATTAACAGAGTCTCCTACTACCGTGTGCTGTGTAACAC 420

Qy 421 CTTTACAGAGTCCAGCTGCTCCCAAGTGGTTCTGTATGAGGCCAAACCCCTGCCAG 480
Db 421 CTTTACAGAGTCCAGCTGCTCCCAAGTGGTTCTGTATGAGGCCAAACCCCTGCCAG 480

Qy 481 AATGGGCTTACTGCTCCCGGATTAAGCGGAGATCCAAAGTTCACTGTGCTGTCCCGAC 540
Db 481 AATGGGCTTACTGCTCCCGGATTAAGCGGAGATCCAAAGTTCACTGTGCTGTCCCGAC 540

Qy 541 CAGTTCAAGGGAAATCTCTGTAATAGTCTGATGACTGCTATGTTGGCGATGGCTAC 600
Db 541 CAGTTCAAGGGAAATCTCTGTAATAGTCTGATGACTGCTATGTTGGCGATGGCTAC 600

Qy 601 TCTTACCGAGGAAATGATAGGACAGTCAACAGCATGCTGCTTCTTACTGGAATCC 660
Db 601 TCTTACCGAGGAAATGATAGGACAGTCAACAGCATGCTGCTTCTTACTGGAATCC 660

Qy 661 CACCTCTCTTGCAGGAGAAATTAACAATGTTTATGGAGGATGCTGAAACCATGGGAT 720
Db 661 CACCTCTCTTGCAGGAGAAATTAACAATGTTTATGGAGGATGCTGAAACCATGGGAT 720

Qy 721 GGGGAACAATTTCTGCAGAAACCCAGATGGGACGAAAGCCCTGCTGTTTATAA 780
Db 721 GGGGAACAATTTCTGCAGAAACCCAGATGGGACGAAAGCCCTGCTGTTTATAA 780

Qy 781 GTTACCAATGACAAGGTGAAATGGGAATACTGTGATGTTCTCAGCCCTGCTCAGCCAGAC 840
Db 781 GTTACCAATGACAAGGTGAAATGGGAATACTGTGATGTTCTCAGCCCTGCTCAGCCAGAC 840

Qy 841 GTTGCTTACCCAGAGGAAAGCCCACTGAGCAATCAACCAAGCTTCCGGGTTTACTCC 900
Db 841 GTTGCTTACCCAGAGGAAAGCCCACTGAGCAATCAACCAAGCTTCCGGGTTTACTCC 900

Qy 901 TGTGGAAGACTGAGTATGAGAGAGGAGATCAAGAGAATCTATGGAGGCTTTAAGAGC 960
Db 901 TGTGGAAGACTGAGTATGAGAGAGGAGATCAAGAGAATCTATGGAGGCTTTAAGAGC 960

Qy 961 ACGGGGGCAAGCACCCTGATGGCAGGCTCCCTCCAGTCTCTGCTGCTCTGACCACTCC 1020
Db 961 ACGGGGGCAAGCACCCTGATGGCAGGCTCCCTCCAGTCTCTGCTGCTCTGACCACTCC 1020

Qy 1021 ATGCCCCAGGGCCACTTCTGTGGTGGCGCTGATCCACCCCTGCTGGGTGCTCACTGCT 1080
Db 1021 ATGCCCCAGGGCCACTTCTGTGGTGGCGCTGATCCACCCCTGCTGGGTGCTCACTGCT 1080

Qy 1081 GCCCACTGACCCGACATAAAAAACAGACATCTAAAGTGGTCTAGGGGACAGGACCTG 1140
Db 1081 GCCCACTGACCCGACATAAAAAACAGACATCTAAAGTGGTCTAGGGGACAGGACCTG 1140

Qy 1141 AAGAAAGAGATTTCTAGCAGCAGAGCTTTAGGGTGCAAGATATTCAAGTACAGCCAC 1200
Db 1141 AAGAAAGAGATTTCTAGCAGCAGAGCTTTAGGGTGCAAGATATTCAAGTACAGCCAC 1200

Qy 1201 TACAATGAAGAGATGAGATTTCCCAATGATATTGCAATGCTCAAGTTAAAGCCAGTG 1260
Db 1201 TACAATGAAGAGATGAGATTTCCCAATGATATTGCAATGCTCAAGTTAAAGCCAGTG 1260

Qy 1261 GATGTCATCTGCTCTAGAAATCCAAATACGTGAAGACTGTGCTGCTGCTGATGGTCC 1320
Db 1261 GATGTCATCTGCTCTAGAAATCCAAATACGTGAAGACTGTGCTGCTGCTGATGGTCC 1320

Qy 1321 TTTCCCTCTGGGAGTGAGTGCCACATCTCTGCTGGGTGTTTACAGAAACAGGAAAGGG 1380
Db 1321 TTTCCCTCTGGGAGTGAGTGCCACATCTCTGCTGGGTGTTTACAGAAACAGGAAAGGG 1380

Qy 1381 TCCCGCCAGCTCTGATGCCAAAGTCAAGCTGATGCCAACACTTTGTGCACTCCCGC 1440
Db 1381 TCCCGCCAGCTCTGATGCCAAAGTCAAGCTGATGCCAACACTTTGTGCACTCCCGC 1440

Qy 1441 CAACTCTATGACACATGATTTGATGACATGATGATGCTGTCAGGAAATCTTCAGAAACCT 1500
Db 1441 CAACTCTATGACACATGATTTGATGACATGATGATGCTGTCAGGAAATCTTCAGAAACCT 1500

Qy 1501 GGGCAAGACACTGTCAGGGTGACTGTGAGGCCCTTGAAGAGGAGGAGGAGGAGGAGG 1560
Db 1501 GGGCAAGACACTGTCAGGGTGACTGTGAGGCCCTTGAAGAGGAGGAGGAGGAGGAGG 1560

Qy 1561 TACTAGCTCTATGGGATGAGTGGGCTGGAGTGTGAGAGGAGGAGGAGGAGGAGGAGG 1620
Db 1561 TACTAGCTCTATGGGATGAGTGGGCTGGAGTGTGAGAGGAGGAGGAGGAGGAGGAGG 1620

Qy 1621 ACCCAAGTTACCAAAATTCCTGAAATCGAATCAAGCCCAATCAAAAGTGAAGTGGCTTC 1680
Db 1621 ACCCAAGTTACCAAAATTCCTGAAATCGAATCAAGCCCAATCAAAAGTGAAGTGGCTTC 1680

```

QY 1681 TAA 1683
DB 1681 TAA 1683

RESULT 5
US-09-912-559-1
; Sequence 1, Application US/09912559
; Patent No. US20020142316A1
; GENERAL INFORMATION:
; APPLICANT: ROEMISCH, JUERGEN
; APPLICANT: STOEHR, HANS-ARNOLD
; APPLICANT: FEUSSNER, ANNETTE
; APPLICANT: LANG, WIEGAND
; APPLICANT: WEINER, THOMAS
; APPLICANT: BECKER, MARGRET
; APPLICANT: NERLICH, CLAUDIA
; APPLICANT: MUTH-NAUMANN, GUDRUN
; TITLE OF INVENTION: MUTANTS OF THE FACTOR VII-ACTIVATING PROTEASE AND
; FILE REFERENCE: 06478.1457
; CURRENT APPLICATION NUMBER: US/09/912,559
; CURRENT FILING DATE: 2001-07-26
; PRIOR APPLICATION NUMBER: DE 100 36 641.4
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: DE 100 50 040.4
; PRIOR FILING DATE: 2000-10-10
; PRIOR APPLICATION NUMBER: DE 100 52 319.6
; PRIOR FILING DATE: 2000-10-21
; PRIOR APPLICATION NUMBER: DE 101 18 706.8
; PRIOR FILING DATE: 2001-04-12
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 1
; LENGTH: 1683
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-912-559-1

Query Match 99.88; Score 1679.8; DB 9; Length 1683;
Best Local Similarity 99.98; Pred. No. 0;
Matches 1681; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATGTTGCCAGGATGCTGATCTCCATGTTCTGCTGTTAATGCTCTGTTGGGAAGACA 60
DB 1 ATGTTGCCAGGATGCTGATCTCCATGTTCTGCTGTTAATGCTCTGTTGGGAAGACA 60

QY 61 GCCTGTGGGTTCTCCCTGATGTCCTTTATGGAAGCCTGGACCCAGACTGGACCCCTGAC 120
DB 61 GCCTGTGGGTTCTCCCTGATGTCCTTTATGGAAGCCTGGACCCAGACTGGACCCCTGAC 120

QY 121 CAGTATGATTACAGCTACGAGGATTATATCAGGAAGAGAACACCACTAGCACACTTACC 180
DB 121 CAGTATGATTACAGCTACGAGGATTATATCAGGAAGAGAACACCACTAGCACACTTACC 180

QY 181 CATGCTGAGATCTCTGACTGCTGACTACACTGAGGACCAAGCTGATCCATGCCAGCCCAAC 240
DB 181 CATGCTGAGATCTCTGACTGCTGACTACACTGAGGACCAAGCTGATCCATGCCAGCCCAAC 240

QY 241 CCCTGTGAACACGGTGGGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 300
DB 241 CCCTGTGAACACGGTGGGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 300

QY 301 GCTCCTTTCTCTGGGAATTAAGTGTGAGAAAGTCAAAATACGTGCAAGGACCAACCCATGT 360
DB 301 GCTCCTTTCTCTGGGAATTAAGTGTGAGAAAGTCAAAATACGTGCAAGGACCAACCCATGT 360

QY 361 GGCGGGGCCAATGCTCATTAACAGAGTCTCTCTACTACCCCTGTGCTGTGTAACAC 420
DB 361 GGCGGGGCCAATGCTCATTAACAGAGTCTCTCTACTACCCCTGTGCTGTGTAACAC 420

QY 421 CCTTACAGAGTCCAGGCTGCTCCCAAGTGTCTGCTGATGCAAGCAAAACCCCTGCCAG 480
DB 421 CCTTACAGAGTCCAGGCTGCTCCCAAGTGTCTGCTGATGCAAGCAAAACCCCTGCCAG 480

```

```

DB 421 CCTTACAGAGTCCAGGCTGCTCCCAAGTGTCTCTGATGCAAGCAAAACCCCTGCCAG 480
QY 481 AATGGGGCTACCTGCTCCGGCATTAAGCGGAGATCCAAAGTTCACTGTGCTCTGCTCCGAC 540
DB 481 AATGGGGCTACCTGCTCCGGCATTAAGCGGAGATCCAAAGTTCACTGTGCTCTGCTCCGAC 540
QY 541 CAGTTCAAGGGGAAATTTCTGTAAGTGTCTGTAAGTGTCTGTAAGTGTCTGTAAGTGTCTG 600
DB 541 CAGTTCAAGGGGAAATTTCTGTAAGTGTCTGTAAGTGTCTGTAAGTGTCTGTAAGTGTCTG 600
QY 601 TCTTACCGAGGAAATTAAGTGTGTAAGTGTGTAAGTGTGTAAGTGTGTAAGTGTGTAAGTGT 660
DB 601 TCTTACCGAGGAAATTAAGTGTGTAAGTGTGTAAGTGTGTAAGTGTGTAAGTGTGTAAGTGT 660
QY 661 CACTCTCTCTGAGGAAATTAACAATGTAAGTGTGTAAGTGTGTAAGTGTGTAAGTGTGTAAG 720
DB 661 CACTCTCTCTGAGGAAATTAACAATGTAAGTGTGTAAGTGTGTAAGTGTGTAAGTGTGTAAG 720
QY 721 GGGGAACACAAATTTCTGCAAGAACCCAGATGCGGAGGAAAGCCCTGCTGCTTTATATAA 780
DB 721 GGGGAACACAAATTTCTGCAAGAACCCAGATGCGGAGGAAAGCCCTGCTGCTTTATATAA 780
QY 781 GTTACCAATCAAAAGTGTGAAATGGAATTAATGGAATTAATGGAATTAATGGAATTAATGGA 840
DB 781 GTTACCAATCAAAAGTGTGAAATGGAATTAATGGAATTAATGGAATTAATGGAATTAATGGA 840
QY 841 GTTGCCTTACCGAGGAAAGCCCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAG 900
DB 841 GTTGCCTTACCGAGGAAAGCCCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAG 900
QY 901 TGTGAAAGACTGAGATGAGAGAGGAAAGATCAAGAGAAATCTATGAGAGCTTTAAGAGC 960
DB 901 TGTGAAAGACTGAGATGAGAGAGGAAAGATCAAGAGAAATCTATGAGAGCTTTAAGAGC 960
QY 961 ACGCGGGCAAGACCCATGCGAGGCTGCTCCCTGAGTCTGCTGCTGCTGCTGCTGCTGCTGCT 1020
DB 961 ACGCGGGCAAGACCCATGCGAGGCTGCTCCCTGAGTCTGCTGCTGCTGCTGCTGCTGCTGCT 1020
QY 1021 ATGCCCCAGGSCCACTTCTGTGTGGGGCGCTGATCCACCCCTGCTGCTGCTGCTGCTGCT 1080
DB 1021 ATGCCCCAGGSCCACTTCTGTGTGGGGCGCTGATCCACCCCTGCTGCTGCTGCTGCTGCT 1080
QY 1081 GCCCCTGCAACCGACATAAAGGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAG 1140
DB 1081 GCCCCTGCAACCGACATAAAGGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAG 1140
QY 1141 AAGAAAGAGAAATTTCTGAGCAGAGCTTTAGGCTGAGAGATATTAAGATCAAGAGCAAC 1200
DB 1141 AAGAAAGAGAAATTTCTGAGCAGAGCTTTAGGCTGAGAGATATTAAGATCAAGAGCAAC 1200
QY 1201 TACAATGAAAGAGATGAGATTTCCCAATGATATTTGATTTGCTCAAGTTAAAGCCAGTG 1260
DB 1201 TACAATGAAAGAGATGAGATTTCCCAATGATATTTGATTTGCTCAAGTTAAAGCCAGTG 1260
QY 1261 GATGCTCACTGCTGTAGATTTCAATACCTGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1320
DB 1261 GATGCTCACTGCTGTAGATTTCAATACCTGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1320
QY 1321 TTTTCCCTCTGGGAGTGAAGTGCACATCTCTGCTGGGGTGTACAGAAACAGGAAAGGG 1380
DB 1321 TTTTCCCTCTGGGAGTGAAGTGCACATCTCTGCTGGGGTGTACAGAAACAGGAAAGGG 1380
QY 1381 TCCCGCAGCTCTCGATGCCAAGTCAAGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1440
DB 1381 TCCCGCAGCTCTCGATGCCAAGTCAAGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1440
QY 1441 CAACTCTATGACCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1500
DB 1441 CAACTCTATGACCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1500
QY 1501 GGGCAAGACACCTGCGAGGCTGCTGAGGCCCCCTGAGCTGCTGAGAGAGAGGAGGAC 1560
DB 1501 GGGCAAGACACCTGCGAGGCTGCTGAGGCCCCCTGAGCTGCTGAGAGAGAGGAGGAC 1560

```

Qy 1561 TACTAGCTCTATGGATAGTACGCTGGGCTGGAGTGTGAGAGAGGAGCCAGGGCTCTAC 1620
Db 1561 TACTAGCTCTATGGATAGTACGCTGGGCTGGAGTGTGAGAGAGGAGCCAGGGCTCTAC 1620
Qy 1621 ACCCAAGTTACCAAAATTCCTGAATGGATCAAAAGCCACCACATCAAAAGTGAAGTGGCTTC 1680
Db 1621 ACCCAAGTTACCAAAATTCCTGAATGGATCAAAAGCCACCACATCAAAAGTGAAGTGGCTTC 1680
Qy 1681 TAA 1683
Db 1681 TAA 1683
RESULT 6
US-10-391-215-1
; Sequence 1, Application US/10391215
; Publication No. US20040009543A1
; GENERAL INFORMATION:
; APPLICANT: KIEHL, STEFAN
; APPLICANT: WILLEIT, JOHANN
; APPLICANT: WIEDERMANN, CHRISTIAN JOSEF
; APPLICANT: ROEMISCH, JUERGEN
; APPLICANT: WEIMER, THOMAS
; APPLICANT: FEUSSNER, ANNETTE
; APPLICANT: STOEHR, HANS-ARNOLD
; APPLICANT: DOERSAM, VOLKER
; TITLE OF INVENTION: (FSAP) I MUTANT OF FACTOR VII ACTIVATING PROTEASE
; TITLE OF INVENTION: (FSAP) AS RISK FACTOR FOR ARTERIAL THROMBOISIS
; FILE REFERENCE: 06478 1457-01
; CURRENT APPLICATION NUMBER: US/10391,215
; CURRENT FILING DATE: 2003-03-19
; PRIOR APPLICATION NUMBER: 09/912,559
; PRIOR FILING DATE: 2001-07-26
; PRIOR APPLICATION NUMBER: DE 100 36 641.4
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: DE 100 50 040.4
; PRIOR FILING DATE: 2000-10-10
; PRIOR APPLICATION NUMBER: DE 100 52 319.6
; PRIOR FILING DATE: 2000-10-21
; PRIOR APPLICATION NUMBER: DE 101 18 706.8
; PRIOR FILING DATE: 2001-04-12
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 1683
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-391-215-1
Query Match 99.8%; Score 1679.8; DB 16; Length 1683;
Best Local Similarity 99.9%; Pred No. 0;
Matches 1681; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 1 ATGTTGCCAGATGTCGTGATCTCATGTTCTGCTGTTAATGGCTCTGGTGGGAAGACA 60
Db 1 ATGTTGCCAGATGTCGTGATCTCATGTTCTGCTGTTAATGGCTCTGGTGGGAAGACA 60
Qy 61 GCCTGTGGTCTCCCTCATGCTCTTATTGAAAGCCCTGGACCAGACTGGACCCCTGAC 120
Db 61 GCCTGTGGTCTCCCTCATGCTCTTATTGAAAGCCCTGGACCAGACTGGACCCCTGAC 120
Qy 121 CAGTATGATTACAGTACGAGATTATTAATCAGGAAGAGAACACAGTACACACTTACC 180
Db 121 CAGTATGATTACAGTACGAGATTATTAATCAGGAAGAGAACACAGTACACACTTACC 180
Qy 181 CATGCTGAGATCCTGACTGTTACTACTAGGAGCCAGCTGATCCATGCGCCGAC 240
Db 181 CATGCTGAGATCCTGACTGTTACTACTAGGAGCCAGCTGATCCATGCGCCGAC 240
Qy 241 CCCTGTGAACACGGTGGGAGCTGCCCTCGTCCATGGGAGCACTTTCACATGAGCTGCTG 300
Db 241 CCCTGTGAACACGGTGGGAGCTGCCCTCGTCCATGGGAGCACTTTCACATGAGCTGCTG 300

Qy 301 GCTCCTTTCTCTGGAAATAAGTGTGAGAAAGTGCAGAAATAGTGCAGAGGACAAACCCATGT 360
Db 301 GCTCCTTTCTCTGGAAATAAGTGTGAGAAAGTGCAGAAATAGTGCAGAGGACAAACCCATGT 360
Qy 361 GCGCGGGGCCAAATGTCTCATTTACCCAGAGTCTCCCTACTACCCCTGTGTGTAAACAC 420
Db 361 GCGCGGGGCCAAATGTCTCATTTACCCAGAGTCTCCCTACTACCCCTGTGTGTAAACAC 420
Qy 421 CTTTACACAGTCCAGCTGCTCCAGTGGTCTCTGTATGACAGGCAAAACCCCTGCCAG 480
Db 421 CTTTACACAGTCCAGCTGCTCCAGTGGTCTCTGTATGACAGGCAAAACCCCTGCCAG 480
Qy 481 AATGGGGCTACCTGCTCCCGCATAAAGCGGAGATCCAAAGTTCACCTGTGCTCTCCGAC 540
Db 481 AATGGGGCTACCTGCTCCCGCATAAAGCGGAGATCCAAAGTTCACCTGTGCTCTCCGAC 540
Qy 541 CAGTTCAAGGGGAAATCTGTGAAATAGTGTCTGATGACTCTCTATGTGGGATGGGTAC 600
Db 541 CAGTTCAAGGGGAAATCTGTGAAATAGTGTCTGATGACTCTCTATGTGGGATGGGTAC 600
Qy 601 TCTTACCGAGGAAATGAATAGGACAGTCAACAGCATGCGTGCCTTTACTTGAACCTCC 660
Db 601 TCTTACCGAGGAAATGAATAGGACAGTCAACAGCATGCGTGCCTTTACTTGAACCTCC 660
Qy 661 CACCTCCTCTTCGAGGAGATTAACAACATGTTTATGGAGGATGCTGAAACCCATGGGAT 720
Db 661 CACCTCCTCTTCGAGGAGATTAACAACATGTTTATGGAGGATGCTGAAACCCATGGGAT 720
Qy 721 GGGGAACACAAATTTCTGCAGAAACCCAGATCGGACGAAAGCCCTGTGTATTATAA 780
Db 721 GGGGAACACAAATTTCTGCAGAAACCCAGATCGGACGAAAGCCCTGTGTATTATAA 780
Qy 781 GTTACCAATGACAGGTGAAATGGGAACTGATGCTGATGCTCAGCTGCTCAGCCAGGAC 840
Db 781 GTTACCAATGACAGGTGAAATGGGAACTGATGCTGATGCTCAGCTGCTCAGCCAGGAC 840
Qy 841 GTTGCTTACCCAGAGGAAAGCCCACTGAGCCATCAACCAAGCTTCGCGGTTGACTCC 900
Db 841 GTTGCTTACCCAGAGGAAAGCCCACTGAGCCATCAACCAAGCTTCGCGGTTGACTCC 900
Qy 901 TGTGAAAGACTGAGATGACAGAGAGGAAAGTCAAGAGAACTCTATGAGGCTTTAAGAGC 960
Db 901 TGTGAAAGACTGAGATGACAGAGAGGAAAGTCAAGAGAACTCTATGAGGCTTTAAGAGC 960
Qy 961 ACGGGGGCAAGCACCCTATGGCAGCGTCCCTCCAGTCTCTGCTGCTCTGACCATCTCC 1020
Db 961 ACGGGGGCAAGCACCCTATGGCAGCGTCCCTCCAGTCTCTGCTGCTCTGACCATCTCC 1020
Qy 1021 ATGCCCCAGGCGCACTTCTGTGGTGGGCGTGTATCCACCCCTGTGGTGTCTACTGCT 1080
Db 1021 ATGCCCCAGGCGCACTTCTGTGGTGGGCGTGTATCCACCCCTGTGGTGTCTACTGCT 1080
Qy 1081 GCCCACTGCACGACATAAAACACAGACATCTAAAGGTGGTGTCTAGGGGACCCAGGACCTG 1140
Db 1081 GCCCACTGCACGACATAAAACACAGACATCTAAAGGTGGTGTCTAGGGGACCCAGGACCTG 1140
Qy 1141 AAGAAAGAAATTTCTAGAGCAGAGCTTTAGGGTGCAGAGATATTCAAGTACAGCCAC 1200
Db 1141 AAGAAAGAAATTTCTAGAGCAGAGCTTTAGGGTGCAGAGATATTCAAGTACAGCCAC 1200
Qy 1201 TACAATGAAGAGATGAGATTTCCACATGATATTGCTCAAGTAAAGCCAGTG 1260
Db 1201 TACAATGAAGAGATGAGATTTCCACATGATATTGCTCAAGTAAAGCCAGTG 1260
Qy 1261 GATGGTCACTGTGCTCTAGAAATCCAAATACGTGAAGACTGTGTGCTGCTGATGGGTCC 1320
Db 1261 GATGGTCACTGTGCTCTAGAAATCCAAATACGTGAAGACTGTGTGCTGCTGATGGGTCC 1320
Qy 1321 TTTCCCTCTGGAGTGAAGTGGCCACATCTCTGGCTGGGTGTTACAGAAACAGAAAGGG 1380
Db 1321 TTTCCCTCTGGAGTGAAGTGGCCACATCTCTGGCTGGGTGTTACAGAAACAGAAAGGG 1380

QY	1381	TCCCGCAGCTCCTCGGATGCCAAAGTCAAAGCTGATTGCCAACACTTTGTGCAACTCCCGC	1444
Db	1381	TCCCGCAGCTCCTGGATGCCAAAGTCAAAGCTGATTGCCAACACTTTGTGCAACTCCCGC	1440
QY	1441	CAACTCTATGACACACATGATTGATGACAGTATGATCTGTGCAGGNAATCTTCAGAAACCT	1500
Db	1441	CAACTCTATGACACACATGATTGATGACAGTATGATCTGTGCAGGAAATCTTCAGAAACCT	1500
QY	1501	GGGCAAGACACCTGTGCCAGGCTGACTCTGGAGGCCCCCTGACCTGTGAGAAGGACGGCACC	1560
Db	1501	GGGCAAGACACCTGTGCCAGGCTGACTCTGGAGGCCCCCTGACCTGTGAGAAGGACGGCACC	1560
QY	1561	TACTACGCTCTATGGATAGTGAAGTCTGGGCGCTGGAGTGTGAGAAGGACGGGCTCTAC	1620
Db	1561	TACTACGCTCTATGGATAGTGAAGTCTGGGCGCTGGAGTGTGGGAAGACGGGCTCTAC	1620
QY	1621	ACCCAAGTTACCAAAATTCCTGAATTTGGATCAAAGCCACCATCAAAGTGAAAGTGGCTTC	1680
Db	1621	ACCCAAGTTACCAAAATTCCTGAATTTGGATCAAAGCCACCATCAAAGTGAAAGTGGCTTC	1680
QY	1681	TAA 1683	
Db	1681	TAA 1683	
RESULT 7			
US-09-880-107-1668			
; Sequence 1668, Application US/09880107			
; Patent No. US20020142981A1			
; GENERAL INFORMATION:			
; APPLICANT: Horne, Darci T.			
; APPLICANT: Vockley, Joseph G.			
; APPLICANT: Scherf, Uwe			
; APPLICANT: Gene Logic, Inc.			
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer			
; FILE REFERENCE: 44921-5028-WO			
; CURRENT APPLICATION NUMBER: US/09/880,107			
; CURRENT FILING DATE: 2001-06-14			
; PRIOR APPLICATION NUMBER: US 60/211,379			
; PRIOR FILING DATE: 2000-06-14			
; PRIOR APPLICATION NUMBER: US 60/237,054			
; PRIOR FILING DATE: 2000-10-02			
; NUMBER OF SEQ ID NOS: 3950			
; SOFTWARE: PatentIn Ver. 2.1			
; SEQ ID NO 1668			
; LENGTH: 3008			
; TYPE: DNA			
; ORGANISM: Homo sapiens			
; FEATURE:			
; OTHER INFORMATION: Genbank Accession No. US20020:42981A1 D49742			
US-09-880-107-1668			
Query Match 99.8%; Score 1679.8; DB 9; Length 3008;			
Best Local Similarity 99.9%; Pred. No. 0;			
Matches 1681; Conservative 0; Mismatches 2; Indels 0; Gaps 0;			
QY	1	ATGTTTGCCAGAGATCTGATCTCCATGTTCTGCTGTTAATGCTCTGTGGGAAAGACA	60
Db	97	ATGTTTGCCAGAGATCTGATCTCCATGTTCTGCTGTTAATGCTCTGTGGGAAAGACA	156
QY	61	GCTGTGGGTTCTCCCTGATGTCCTTATTGAAAGCTGGACCCAGACTGGACCCCTGAC	120
Db	157	GCTGTGGGTTCTCCCTGATGTCCTTATTGAAAGCTGGACCCAGACTGGACCCCTGAC	216
QY	121	CAGTATGATTACAGCTACAGGATTATTAATCAGGAAGAACACCAGTAGACACTTACC	180
Db	217	CAGTATGATTACAGCTACAGGATTATTAATCAGGAAGAACACCAGTAGACACTTACC	276
QY	181	CATGCTGAGAATCTGACTGTTACTACACTGAGAACCAAGTGTCCATGCCAGCCCAAC	240
Db	277	CATGCTGAGAATCTGACTGTTACTACACTGAGAACCAAGTGTCCATGCCAGCCCAAC	336
QY	241	CCCTGTGAACAGGTTGGGACTGCGCTGCTCATGGGAGCACTTTCATATGCAGCTCGCTG	300

Db 1417 TTTCCCTCTGGAGTGAGTCCACATCTCTGGCTGGGGTGTATACAGAAACAGGAAAGGG 1476
Qy 1381 TCCCGCAGCTCTCTGGATGCCAAGTCAAGCTGATTCGCAACACTTTGTGCACTCCCGC 1440
Db 1477 TCCCGCAGCTCTCTGGATGCCAAGTCAAGCTGATTCGCAACACTTTGTGCACTCCCGC 1536
Qy 1441 CAACCTATGACACACATGATTCGATGACAGTATGATCTGTGCAAGAAATCTTCAGAAACCT 1500
Db 1537 CAACCTATGACACATGATTCGATGACAGTATGATCTGTGCAAGAAATCTTCAGAAACCT 1596
Qy 1501 GGGCAAGACACTCCAGGCTGACTCTGGAGGCCCTCTGACCTGTGAGAGGACGACCC 1560
Db 1597 GGGCAAGACACTCCAGGCTGACTCTGGAGGCCCTCTGACCTGTGAGAGGACGACCC 1656
Qy 1561 TACTACCTCTATGGATAGTGAAGTGGGCTGAGTGTGAGAGAGGCCAGGGCTCTAC 1620
Db 1657 TACTACCTCTATGGATAGTGAAGTGGGCTGAGTGTGAGAGAGGCCAGGGCTCTAC 1716
Qy 1621 ACCCAAGTTACCAATCTCTGAAATGGATCAAGCCACCATCAAAAGTGAAGTGGCTTC 1680
Db 1717 ACCCAAGTTACCAATCTCTGAAATGGATCAAGCCACCATCAAAAGTGAAGTGGCTTC 1776
Qy 1681 TAA 1683
Db 1777 TAA 1779

RESULT 8
US-09-918-995-8242
; Sequence 8242, Application US/0918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 8242
; LENGTH: 428
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-918-995-8242

Query Match 21.9%; Score 368.4; DB 10; Length 428;
Best Local Similarity 99.7%; Pred. No. 4.1e-111;
Matches 369; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 ATGTTTGCAGGATGCTGATCTCCATGTTCTGCTGTTAATGGCTCTGTGGGAAGACA 60
Db 59 ATGTTTGCAGGATGCTGATCTCCATGTTCTGCTGTTAATGGCTCTGTGGGAAGACA 118
Qy 61 GCCTGTGGGTTCCTCCTGATGCTTTATTTGAAAGCCTGGACCCAGACTGGACCCCTGAC 120
Db 119 GCCTGTGGGTTCCTCCTGATGCTTTATTTGAAAGCCTGGACCCAGACTGGACCCCTGAC 178
Qy 121 CAGTATGATACAGCTACAGAGTATATATCAGAGAGAAACACAGTAGCACTTACC 180
Db 179 CAGTATGATACAGCTACAGAGTATATATCAGAGAGAAACACAGTAGCACTTACC 238
Qy 181 CATGCTGAGAATCCTGACTGGTACTACACTGAGGACCAAGCTGATCCATGCCAGCCCAAC 240
Db 239 CATGCTGAGATCCTGACTGTTACTACTAGGACCAAGCTGATCCATGCCAGCCCAAC 298
Qy 241 CCTGTGAACACGGTGGGACTCCCTGCTTCATGAGGACACTTACATGAGCTGCCTG 300
Db 299 CCTGTGAACACGGTGGGACTCCCTGCTTCATGAGGACACTTACATGAGCTGCCTG 358
Qy 301 GCTCCCTTCTCTGGGAATAGTCTCAGAAAGTCAAAATAGCTGCAAGGACCAACCATGT 360

Db 359 GCTCCTTTCTCGGAATAAGTGTGCAAAATGCAAAATAGTGTGCAAGGACACCCATGT 418
Qy 361 GCGCGGGGCC 370
Db 419 GCGCGGGGCC 428
RESULT 9
US-10-425-000-17
; Sequence 17, Application US/10425000
; Publication No. US20040052777A1
; GENERAL INFORMATION:
; APPLICANT: Nesbit, Mark
; APPLICANT: Cameron, Beatrice
; APPLICANT: Blanche, Francis
; TITLE OF INVENTION: Kringlike Polypeptides and Methods for Using Them to Inhibit
; TITLE OF INVENTION: Angiogenesis
; FILE REFERENCE: ST01027-B
; CURRENT APPLICATION NUMBER: US/10/425,000
; CURRENT FILING DATE: 2003-04-29
; PRIOR APPLICATION NUMBER: 10/233,675
; PRIOR FILING DATE: 2002-09-04
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 17
; LENGTH: 264
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Nucleotide sequence encoding the kringlike domain of the hyaluronan
; OTHER INFORMATION: binding protein
US-10-425-000-17

Query Match 15.6%; Score 262.4; DB 13; Length 264;
Best Local Similarity 99.6%; Pred. No. 5e-76;
Matches 263; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 574 GATGACTGTATGTTGGCGATGCTACTCTTACCGAGGAAAATGAATAGGACAGTCAAC 533
Db 1 GATGACTGTATGTTGGCGATGCTACTCTTACCGAGGAAAATGAATAGGACAGTCAAC 60
Qy 634 CAGCATGCGTGCCTTTACTGGAACTCCACCTCTCTTCAGAGAGAAATACAACTGTTT 693
Db 61 CAGCATGCGTGCCTTTACTGGAACTCCACCTCTCTTCAGAGAGAAATACAACTGTTT 120
Qy 694 ATGGAGGATGCTGAAACCCCATGGATGGGGAACACAAATTTCTGCAGAAACCCAGATGCG 753
Db 121 ATGGAGGATGCTGAAACCCCATGGATGGGGAACACAAATTTCTGCAGAAACCCAGATGCG 180
Qy 754 GACGAAAGCCCTGGTGTCTTTAATTAAGTTACCAATGACAGGTGAATGGGTAATCTGT 813
Db 181 GACGAAAGCCCTGGTGTCTTTAATTAAGTTACCAATGACAGGTGAATGGGTAATCTGT 240
Qy 814 GATGCTCAGCCTGCTCAGCCAG 837
Db 241 GATGCTCAGCCTGCTCAGCCTAG 264

RESULT 10
US-09-864-761-11164/c
; Sequence 11164, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Acomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23

;; PRIOR APPLICATION NUMBER: US 60/180,312
;; PRIOR FILING DATE: 2000-02-04
;; PRIOR APPLICATION NUMBER: US 60/207,456
;; PRIOR FILING DATE: 2000-05-26
;; PRIOR APPLICATION NUMBER: US 09/632,366
;; PRIOR FILING DATE: 2000-08-03
;; PRIOR APPLICATION NUMBER: GB 24263.6
;; PRIOR FILING DATE: 2000-10-04
;; PRIOR APPLICATION NUMBER: US 60/236,359
;; PRIOR FILING DATE: 2000-09-27
;; PRIOR APPLICATION NUMBER: PCT/US01/00666
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00667
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00664
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00669
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00665
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00668
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00663
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00662
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00661
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00670
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: US 60/234,687
;; PRIOR FILING DATE: 2000-09-21
;; PRIOR APPLICATION NUMBER: US 09/608,408
;; PRIOR FILING DATE: 2000-06-30
;; PRIOR APPLICATION NUMBER: US 09/774,203
;; PRIOR FILING DATE: 2001-01-29
;; NUMBER OF SEQ ID NOS: 49117
;; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
;; SEQ ID NO 11164
;; LENGTH: 451
;; TYPE: DNA
;; ORGANISM: Homo sapiens
;; FEATURE:
;; OTHER INFORMATION: MAP TO AC006097.1
;; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.89
;; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1
;; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.8
;; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.1
;; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1
;; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.9
;; OTHER INFORMATION: NT HIT: S83182.1, EVALUATE 8.00e-60
;; OTHER INFORMATION: SWISSPROT HIT: Q25464, EVALUATE 1.00e-03
;; OTHER INFORMATION: EST_HUMAN HIT: T68666.1, EVALUATE 7.00e-55
US-09-864-761-11164

Query Match 7.1%; Score 120.2; DB 9; Length 451;
Best Local Similarity 97.6%; Pred. No. 1.2e-28;
Matches 122; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Qy 324 TCAGAAAGTCAAAATACGTGCAAGGACCAACCCATGTGCGCGGGCCAAATGCTCTATTAC 393
Db 451 TCCTACAGTGCAGAAATACGTGCAAGGACCAACCCATGTGCGCGGGCCAAATGCTCTATTAC 392
Qy 384 CCAGAGTCCCTCCCTACTACCGCTGTGCTGTAAACACCCCTTACACAGGTCCTCCAGCTGCTC 443
Db 391 CCAGAGTCCCTCCCTACTACCGCTGTGCTGTAAACACCCCTTACACAGGTCCTCCAGCTGCTC 332
Qy 444 CCAAG 448
Db 331 CCAAG 327

RESULT 11
US-09-864-761-27791/c
; Sequence 27791, Application US/09864761
; Patent No. US20020048763A1

;; GENERAL INFORMATION:
;; APPLICANT: Penn, Sharron G.
;; APPLICANT: Rank, David R.
;; APPLICANT: Hanzel, David K.
;; APPLICANT: Chen, Wensheng
;; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
;; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
;; FILE REFERENCE: Reomica-X-1
;; CURRENT APPLICATION NUMBER: US 09/864,761
;; CURRENT FILING DATE: 2001-05-23
;; PRIOR APPLICATION NUMBER: US 60/180,312
;; PRIOR FILING DATE: 2000-02-04
;; PRIOR APPLICATION NUMBER: US 60/207,456
;; PRIOR FILING DATE: 2000-05-26
;; PRIOR APPLICATION NUMBER: US 09/632,366
;; PRIOR FILING DATE: 2000-08-03
;; PRIOR APPLICATION NUMBER: GB 24263.6
;; PRIOR FILING DATE: 2000-10-04
;; PRIOR APPLICATION NUMBER: US 60/236,359
;; PRIOR FILING DATE: 2000-09-27
;; PRIOR APPLICATION NUMBER: PCT/US01/00666
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00667
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00664
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00669
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00665
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00668
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00663
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00662
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00661
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00670
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: US 60/234,687
;; PRIOR FILING DATE: 2000-09-21
;; PRIOR APPLICATION NUMBER: US 09/608,408
;; PRIOR FILING DATE: 2000-06-30
;; PRIOR APPLICATION NUMBER: US 09/774,203
;; PRIOR FILING DATE: 2001-01-29
;; NUMBER OF SEQ ID NOS: 49117
;; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
;; SEQ ID NO 27791
;; LENGTH: 117
;; TYPE: DNA
;; ORGANISM: Homo sapiens
;; FEATURE:
;; OTHER INFORMATION: MAP TO AC006097.1
;; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.89
;; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1
;; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.8
;; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.1
;; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1
;; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.9
;; OTHER INFORMATION: NT HIT: S83182.1, EVALUATE 8.00e-60
;; OTHER INFORMATION: SWISSPROT HIT: Q25464, EVALUATE 1.00e-03
;; OTHER INFORMATION: EST_HUMAN HIT: T68666.1, EVALUATE 7.00e-55
US-09-864-761-27791

Query Match 7.0%; Score 117; DB 9; Length 117;
Best Local Similarity 100.0%; Pred. No. 6.2e-28;
Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 332 TGCAAAATACGTGCAAGGACCAACCCATGTGCGCGGGCCAAATGCTCTATTACCCAGAGTC 391
Db 117 TGCAAAATACGTGCAAGGACCAACCCATGTGCGCGGGCCAAATGCTCTATTACCCAGAGTC 58

392 CTCCTACTACCGTGTGTTGTAACAACACCCCTTACACAGTGTCACAGTGTCCTCCAAG 448
| | | | |
Db 57 CTCCTACTACCGTGTGTTGTAACAACACCCCTTACACAGTGTCACAGTGTCCTCCAAG 1

RESULT 12
US-09-954-456-552
; Sequence 5S2, Application US/09954456
; Patent No. US20020115057A1
; GENERAL INFORMATION:
; APPLICANT: Young, Paul
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using Can
; TITLE OF INVENTION: Sets
; FILE REFERENCE: 589290-76
; CURRENT APPLICATION NUMBER: US/09/954,456
; CURRENT FILING DATE: 2001-09-18
; PRIOR APPLICATION NUMBER: US/60/233,617
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: US/60/234,052
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: US/60/234,923
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,134
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,637
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US/60/235,638
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US/60/235,711
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,720
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,840
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,863
; PRIOR FILING DATE: 2000-09-27
; NUMBER OF SEQ ID NOS: 2276
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 552
; LENGTH: 2036
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-954-456-552

Query March 6.8%; Score 114.4; DB 9; Length 2036;
Best Local Similarity 49.6%; Pred. No. 2.5e-26;
Matches 605; Conservative 0; Mismatches 561; Indels 54; Gaps 10;

Qy 469 AACCCCTGCAGATGGGGTACTCTGCTCCGGCATGAACGGAGATCCAGTTCACTGT 528
| | | | |
Db 745 AGCCCTTGGCTGAACGGGGGCACCTGCCACTGATCGTGCCACCGGACCAACCGTGT 804
| | | | |
Qy 529 GCCTGTCCCACCAGATTCAAGGGGAAATTCGTGAAATAGGTTCTGATGA---CTGCTAT 585
| | | | |
Db 805 GCCTGCCACCAGGCTTCGCTGGACGGCTCTGCAACATCGAGCCTGATGAGCGCTGCTTC 864
| | | | |
Qy 586 GTTGGGATGGTACTCTTACCGAGGGAAAATAAATAGGACAGTCAACACAGATCGGTGC 645
| | | | |
Db 865 TTGGGGAACGGCACATGGGTATCCGCTGGCGTGCCGCCACCTTCAGCCTCGGSCCTCAGCTGC 924
| | | | |
Qy 646 CTTTACTGGAACTCCCACTCTTTCAGAGAATTTACAACATGTTTATGGAGGATGCT 705
| | | | |
Db 925 CTGGCTTGGAACTCCGATCTGCTCTACAGAGACTGCACGTGGACTCCGTGGGCGCGCG 984
| | | | |
Qy 706 GAACACCCATGGGATTTGGGGAACAATTTCTGCAGAACCCAGATCGGACGAAAAAGCCC 765
| | | | |
Db 985 GCCTGTGTGGCGCTGGGGCCCCCATGCTACTTGCCGGAATCCGGACAATGACGAGAGGCC 1044
| | | | |
Qy 766 TGTGTCCTTTAATAAGTTTACCATGACAGGTGAAATGGGAATCTGTGATGTCCTCAGCC 825
| | | | |
Db 1045 TGTGTGTACGT---GGTGAAGGACAGCGCGCTCTCTCTGGGAGTACTGCCCGCTTGGAGGCC 1101
| | | | |
Qy 826 TGCTCAGCCCGAGGACGTGTGCTATCCCGAGAGAAAGCCCACTGAGGCATCAACCAAGCTT 885
| | | | |

RESULT 13

```

RES011 13
US-09-880-107-1512
; Sequence 1612, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherf, Uwe
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 44921-5028-WO
; CURRENT APPLICATION NUMBER: US/09/880.107
; CURRENT FILING DATE: 2001-06-14

```


; PRIOR APPLICATION NUMBER: US 60/211,379
 ; PRIOR FILING DATE: 2000-06-14
 ; PRIOR APPLICATION NUMBER: US 60/237,054
 ; PRIOR FILING DATE: 2000-10-02
 ; NUMBER OF SEQ ID NOS: 3950
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 1612
 ; LENGTH: 2036
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; OTHER INFORMATION: Genbank Accession No. US20020142981A1 D14012
 ; US-09-880-107-1612

Query Match 6.8%; Score 114.4; DB 9; Length 2036;
 Best Local Similarity 49.6%; Pred. No. 2.5e-26;
 Matches 605; Conservative 0; Mismatches 561; Indels 54; Gaps 10;
 QY 469 AACCCCTGCGAGATGGGGTACCTGCTCCGGCATAACGGAGATCCAGTTCACTGT 528
 DB 745 AGCCCTTGCTGAACGGGGGACCTGCCACCTGATCGTGGCCACCGGACCACTGTGT 804
 QY 529 GCCTGCTCCGACACAGTTCAAGGGGAAATCTGTGAATAGTTCTGATGA---CTGCTAT 585
 DB 805 GCCTGCCACCAAGCTTCGCTGACGGCTCTGCAACATCAGCCCTGATGAGCGCTGCTTC 864
 QY 586 GTTGGCGATGGCTACTCTTACCGAGGGAATGAATAGACAGTCAACCGAGATCGGTGC 645
 DB 865 TTGGGAACCGCACTGGTACCGTGGCGAGCACCTCAGCCCTCGGGCCCTCAGCTGC 924
 QY 646 CTTTACTGGAATCCACCTCTCTTTCAGAGGAAATTACAACATGTTTATGAGGATGCT 705
 DB 925 CTGGCTTGNACTCCGATCTGTCTACAGGAGCTGCAGTGGACTCGGTGGGCCCGGG 984
 QY 706 GAAACCCATGGGATTTGGGAACCAATTTCTGAGAAACCCAGATCGGAGCAAAAGGCC 765
 DB 985 GCCTGTGGGCTGGGGCCCTTCTGTCTGCGGAATCCGGAATCCGCAATCAGGAGGCC 1044
 QY 766 TGTGCTTTTAAAGTTACCAATGACAGGTGAATGGAATCTGATGATCTCTCAGCC 825
 DB 1045 TGTGCTACGT---GGTGAAGACAGGGCTCTCTCTGGAGTACTGCGCCCTGGAGGCC 1101
 QY 826 TGTCTAGCCAGACGTTGCTTACCAGAGGAAAGCCCACTGAGCCATCAACCAAGCTT 885
 DB 1102 TGGGAATCCCT-----CACCAGAGTCCAACTGTCCCGGATCTCTCTGGCGACCTG 1152
 QY 886 CCGGGTTTACCTCTGTGGAAGACTGAGATGACAGAGGAAGATCAAGAGATCTAT 945
 DB 1153 CTGAGCCAGCTCCCGGGGCGCCAGCGCTGCGGAGGAGGCAAGAGAGGAGCTTC 1212
 QY 946 GGAGGCTTTAAGAGCAGCGCGGCAAGCAACCCATGCGAGCGTCTCCAGTCTCTGCTG 1005
 DB 1213 CTGCGGCCAGTATCATCGCGGCTCTCTCTGCTGCGGCTCGCAACCCCTG----- 1265
 QY 1006 CCTCTGACCATCTCCATGCCCCAGGGCCACTTCTGTGGGGCGTGTATCCACCCCTGC 1065
 DB 1266 --GCTGGCCGCCATCTACATCGGGGACAGCTTCTGCGCGGGAGCGCTGTCTCACACCTGC 1323
 QY 1066 TGGGTGCTACTGTCGCCACTGCAACGACATAAACACAGACATCTAA-----AGGTG 1119
 DB 1324 TGGGTGTGTGCGCGGCCCACTGCTCTCCACAGCCCCCCCCAGGACAGCGTCTCCGTG 1383
 QY 1120 GTCTAGGAGACAGAGCTGAGAAAGAAAGAAATTTATGAGAGAGCTTTAGGTGTCAG 1179
 DB 1384 GTCTGGGCGAGCACTTTCTTCAACCGGACGACGAGCGTGAACGAGACCTTCGCGATCGAG 1443
 QY 1180 AAGATATTCAAGTACAGCCACTACAATGAAGAGATGAGATTTCCCAACATATATGCA 1239
 DB 1444 AAGTACATCCGTACACCTGTACTCGGTGTTCAACCCAG---CGACCCAGCCTCTGTC 1500
 QY 1240 TTGCTCAAGTTAAGCCAGTGGATGTCTACTGTCTCTAGATCCAAATACGTGAAGACT 1299
 DB 1501 CTGATCCGGCTGAGAGAAAGGGGACCGCTGTGCCACACGCTCGCAGTTCGTGACGCC 1560

QY 1300 GTGTGCTTGGCTGATG-----GGTCTTTTCCCTCTGGAGTGAGTGCACATCTCTGGC 1353
 DB 1561 ATCTGCTTGGCCGAGCGCGGACACCTTCCCGCAGGACACAAAGTCCAGATTGGGGC 1620
 QY 1354 TGGGG-----TGTACAGAAACAGAAAGGGTCCCGCAGCTCTCTGATGCCAAAGTC 1407
 DB 1621 TGGGGCCCACTTGGATGAGAACGTGAGCGGCTACTCCAGCTCCCTGCGGAGGCCCTGGTC 1680
 QY 1408 AAGCTGATTGCCCAACACTTTGTGCAACTCCCGCAACTCTATGACCAATGATTGATGAC 1467
 DB 1581 CCCCTGCTGCCGACCAACAGTGCAGAGCCCTGAGGTCTACGCGCCGACATCAGCCCC 1740
 QY 1468 AGTATGATCTGTGAGAAATCTTCAGAAACCTGGGCAAGACACCTGCGAGGGTGACTCT 1527
 DB 1741 AACATGCTGTGTCGGCTACTT---CGACTGCAAGTCCGACGCTGCCAGGGGGACTCA 1797
 QY 1528 GGAGGGCCCTGACCTGTGAGAGGAGCGCACCTACTAGTCTCTATGGGATGAGTGTG 1587
 DB 1798 GGGGGGGCCCTGGCTTCGAGAAAGACGGCTGGCTTACTCTTACGGCATCATCAGCTGG 1857
 QY 1588 GGCTGTGAGTGTGAG-----AAGAGGCCAGGGGTCTACACCCAAAGTTACCAAATTCCTG 1641
 DB 1858 GGTGACGGCTGCGGGCGGCTCCCAAGCCGGGGTCTACACCGCGTGGCCAACTATGTG 1917
 QY 1642 AATTGGATCAAGCCACCAT 1661
 DB 1918 GACTGATCAACGACCGAT 1937

RESULT 14

US-10-087-192-590
 ; Sequence 590, Application US/10087192
 ; Publication No. US20020182586A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Morris, David W.
 ; APPLICANT: Engelhard, Eric K.
 ; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
 ; FILE REFERENCE: 529452000122
 ; CURRENT APPLICATION NUMBER: US/10/087,192
 ; CURRENT FILING DATE: 2002-03-01
 ; PRIOR APPLICATION NUMBER: US 09/747,377
 ; PRIOR FILING DATE: 2000-12-22
 ; PRIOR APPLICATION NUMBER: US 09/798,586
 ; PRIOR FILING DATE: 2001-03-02
 ; NUMBER OF SEQ ID NOS: 2059
 ; SOFTWARE: PastSeq for Windows Version 4.0
 ; SEQ ID NO 590
 ; TYPE: DNA
 ; ORGANISM: Mus musculus
 ; US-10-087-192-590

Query Match 5.5%; Score 93.2; DB 13; Length 1302;
 Best Local Similarity 49.7%; Pred. No. 2.1e-19;
 Matches 586; Conservative 0; Mismatches 523; Indels 69; Gaps 11;

QY 526 TGTGCTGTCCCGACCAAGTTCAAGGGGAAATTTCTGAAATAGT---TCTGATGATGC 582
 DB 154 TGCAGCTGCCCAAGGAAATTCAGGGGAGCACTGTGAGATAGATGATCAAAAACCTGC 213
 QY 583 TATGTTGGCATGGCTACTCTTACCGAGGAAATGAATAGGACAGTCAACCCAGCATGCG 642
 DB 214 TATCATGAAATGTGTACTCTTACCGAGGAAAGCCAACTGATACCAAGTGGGCC 273
 QY 643 TGCCTTTACTGGAATCCCACTCTCTTGCAGGAAATTTACAACATGTTTATGGAGAT 702
 DB 274 TGCCTGGCTGGAATGCGCTGTGTCTTTCAGAAACCTTACAATGCCACAGACCTGAT 333
 QY 703 GCTGAACCCATGGATGGGAGACAAATTTCTGCAAGAAACCCAGATCGGCGGAAAG 762
 DB 334 GCTATTAGCTAGGCTGGGGGAAACAAATTTCTGCAAGAAACCCCTGCAACCAAGCGCA 393

763 CCCTGGTCTTTATTAAGTTACCAATGACAAAGTGAATGGAATACTGTGATGTCTCA 822
Db |||||
394 CCCTGGTCTATGTGCAGATTGCCTAAGCAGTT-----TGCCAA 435
Qy |||||
823 GCCTGCTCAGCCAGGACGTTGCTACCCAGAGGAAAGCCCACTGAGCCATCAACCAAG 882
Db |||||
436 GAATGCATGTGCATGAC--TGCTCTCTTAGCAAAAAGCCTTCTCGTCTGTAGACCAAC 493
Qy |||||
883 CTTCGGGGTTGACTCTCTGTGMAAGACTGAGATAGCAGAGGAGAGATCAAGAAATC 942
Db |||||
494 AAGCTTCCAGTGTGGCC-----AGAAGGCTCTAAGGCCCGCGCTTTAAGATT 540
Qy |||||
943 TATGAGGCTTTAAGACAGCGGGCAAGCACCATTGGCAGGCGTCCCTCCAGTCTCTG 1002
Db |||||
541 GTTGGGGAGAAATTCATGAGTGTGAGAACACAGCCCTGGTTCCGAGCCAT-----CTAC 594
Qy |||||
1003 CTGCTCTGACCATCTCCATGCCCCAGGGCCACTTCTGTGTGGGGGGCTGTATCCACCC 1062
Db |||||
595 CAGAAGAACACAGGAGGAAGTCTCTCCCTCTTAAATGTGTGGGAGTCTCATCAGTCT 654
Qy |||||
1063 TGCTGGTGTCTCACTGCTGCCACTGCAAC-----GACATAAAACACGACATCTAAAG 1116
Db |||||
655 TGCTGGTGGCCAGTGGCCACACTGCTTCAATCACTCCAAAGAGAAACTACGTT 714
Qy |||||
1117 GTGTGCTAGGGACACGAGCTCTGAAGAAAGAAATTTATGAGCAGAGCTTTAGGGTG 1176
Db |||||
715 GTCTACCTGGTCACTCGAAGAGAGCTCTATAATCTGTGAGAGATGAATTTGAGGTG 774
Qy |||||
1177 CAGAGATATTCAGTACAGCCACTACATGAAGAGATGAGATCCCAACATGATATT 1236
Db |||||
775 GAGCAGCTCATCTTGCACGAATATCTACAGGAGAGAGCTGGCTTACCATAATGATATT 834
Qy |||||
1237 GCATTGTCTCAAGTTAAAGCCAGTGGTCTCTAGAAATCCAAATACGTTGAAG 1296
Db |||||
835 GCCTTGTGAAGATACGTACACAGCGGGCAATGTGCACAGCCATCCAGTCCATACAG 894
Qy |||||
1297 ACTGTGTCTTGCCTGATGGTCTT-----TCCCTCTGGAGTGTAGTCCCATCTCT 1350
Db |||||
895 ACCATCTGCTGCCCCCAAGTTTACTGATGCTCCGTTTGGTTTCCAGACTGTGAGTCACT 954
Qy |||||
1351 GCCTGGGTG--GTTACAGAAACAGGAAAGGGTCCCGCCAGCTCTCTGGA---TGCCAAA 1404
Db |||||
955 GCTTTGAAAGAGTCTGAAAGTGACTATCTCTATCCAAAGAACCTGMAAATGTCGTC 1014
Qy |||||
1405 GTCAAGCTGATGCAACACATTTGTGTGMACTCCCGCCCACTCTATGACCATGATGAT 1464
Db |||||
1015 GTAAAGCTTGTCTTCATGAACAGTGTATGCAGGCCCACTACTATGGCTCTGAAATTAAT 1074
Qy |||||
1465 GACAGTATGATCTGTGCAGGAAATCTTCAGAAACCTGGGCAAGACACCTGCCAGGTGAC 1524
Db |||||
1075 TATAAATGCTGTGTGCTGCGGACCCAGGTGGAAC--AGATTCTGCAAGGGCGAT 1131
Qy |||||
1525 TCTGAGGCCCTGACCTGTGAGAGGACCGGCACTACTAGTCTATGGAATAGTGAGC 1584
Db |||||
1132 TCTGAGGACCGCTTATCTGTAACATCGAAGCGGCCCACTCTGAGTGGGATTTGAGC 1191
Qy |||||
1585 TGGGCCCTGGAGTGTG-----AGAGAGGCCGCGGTCTACCCCAAGTTTACCAATTC 1638
Db |||||
1192 TGGGCCCGAGGATGTGAGAGAAACAAAGCCCGGTGTCTACAGAGGCTCTCACACTTC 1251
Qy |||||
1639 CTGAATTGGATCAAGCCCACTCAATCAAAAGTGAAGTGG 1676
Db |||||
1252 CTGGACTGGATTCAATCCACATCTGGAGAGAGAAAG 1289

RESULT 15

US-10-665-216-10

; Sequence 10, Application US/10665216

; Publication No. US20040043957A1

; GENERAL INFORMATION:

; APPLICANT: Brenda F. Baker

; APPLICANT: Susan M. Freier

APPLICANT: Andrew T. Watt
TITLE OF INVENTION: ANTISENSE MODULATION OF UROKINASE PLASMINOGEN ACTIVATOR EXPRESSIO
FILE REFERENCE: RUS-0188
CURRENT APPLICATION NUMBER: US/10/665,216
CURRENT FILING DATE: 2003-09-19
PRIOR APPLICATION NUMBER: US/09/821,972
PRIOR FILING DATE: 2001-03-30
NUMBER OF SEQ ID NOS: 168
SEQ ID NO 10
LENGTH: 2299
TYPE: DNA
ORGANISM: Mus musculus
FEATURE:
NAME/KEY: CDS
LOCATION: (59)...(1360)
US-10-665-216-10

Query Match

Best Local Similarity 5.5%; Score 93.2; DB 13; Length 2299;

Matches 586; Conservative 0; Mismatches 523; Indels 69; Gaps 11;

Qy 526 TGTGCTGTCTCCGACCAAGTTCAGAGGGGAAATTTCTGTGAAATAGGT---TCTGATGACTGC 582
Db 212 TGCAGCTGCCCAAGAAATTCAGGGGAGCACTGTGAGATAGATGCATCAAAAACCTGC 271
Qy 583 TATGTTGGGATGGCTACTCTTACCAGGGAATATGATAGCAGTCAACAGCATGG 642
Db 272 TATCATGGAATGGTGTACTCTTACCAGGGAAGGCCAACACTGATACCAGAGTGGGCC 331
Qy 643 TGCCTTTACTGAACTCCACCTCTCTTGCAGGAGAAATTACAACTTTTATGGAGAT 702
Db 332 TGCCTGGCTGGAATGCCCTGTCTCTTCAAGAACCTCAATGCCACAGACTGAT 391
Qy 703 GTGAAACCCATGGGATTTGGGGAACACAAATTTCTGCAGAAACCCAGATGCGGACGAAAG 762
Db 392 GCTATTAGCTAGGCTGGGGAACACAAATTTACTGCAGAAACCCCTGACCAACAGAGCGA 451
Qy 763 CCCTGGTCTTTATTAAGTTTACCAATGACAGGTGAAATGGAATACTGTGATGTCTCA 822
Db 452 CCCTGGTCTATGTGCAGATTGGCCCTAAGGCAGTT-----TGTCAA 493
Qy 823 GCCTGCTAGCCAGCAGCTTCCCTACCCAGAGGAAAGCCCACTGAGCCATCAACCAAG 882
Db 494 GAATGATGCTGCATGAC--TGCTCTCTTAGCAAAAAGCCTTCTTCGTCTGTAGACCAAC 551
Qy 883 CTTCGGGGTTTGACTCTCTGTGAAAGACTGAGATAGCAGAGAGGAAGATCAAGAGAATC 942
Db 552 AAGGCTTCCAGTGTGGCC-----AGAAGGCTCTAAGGCCCGCTTTAAGATT 598
Qy 943 TATGAGGCTTTAAGAGCAGCGCGGCAAGCACCCATGCGAGGCGTCCCTCCAGTCTCG 1002
Db 599 GTTGGGGAGAATTCATGAGGTGGAGAACCAGCCCTGGTTCGAGCCAT-----CTAC 652
Qy 1003 CTGCTCTGACCATCTCCATGCCCCAGGGCCACTTCTGTGTGGGGCGCTGATCCACCC 1062
Db 653 CAGAAGAACAGGGAGGAAGTCTCTCCCTCTTTAAATGTGTGGGAGTCTCATCAGTCT 712
Qy 1063 TGCTGGTGTCTCACTGCTGCCACTGCACC-----GACATAAAACCCAGACATCTAAAG 1116
Db 713 TGCTGGTGGCAGTGGCCGACACACTGCTTCAATCAACTCCCAAGAGAAACTACGTT 772
Qy 1117 GTGGTCTAGGGGACAGGACCTGAAAGAAAGAAATTTTCATGAGCAGAGCTTTAGGGTG 1176
Db 773 GTCTACCTGGGTCACTGGAAGGAGAGTCTCTTAATCTCTGGAGAGATGAAGTTTGGGGTG 832
Qy 1177 CAGAAGATATTCAAGTACAGCCACTACAATGAAGAGATGAGATTCCTCCCAATGATATT 1236
Db 833 GAGCAGCTCATCTTGCACGNACTACTACAGGGAAGACGCTGGCTTACCATAATGATATT 892
Qy 1237 GCATTGCTCAAGTTAAAGCCAGTGGATGTGCTACTGTGCTCTAGAAATCCAAATACGTTGAAG 1296
Db 893 GCCTTGTGAAGATAGTATCCAGCAGGGGCCAATGTGACAGCCATCCAGGTCCATACAG 952

QY	1297	ACTGTGTGCTTCCCTGATGGTCCCTT-----TCCTCTGGAGTGAGTGCCACATCTCT	1350
Db	953	ACCATCTGCTGCCCCCAAGTTTACTGATGCTCCGTTTGGTTTCAGACTGTGAGATCACT	1012
QY	1351	GGCTGGGT--GTTACAGAAACAGGAAAGGTCCTCCGACGCTCCTGGA---TGCCAAA	1404
Db	1013	GGCTTTGGAAGAGTCTGAAAGTGAATCTCTATCCAAAGAACTGAAATGTCCTC	1072
QY	1405	GTCAGCTGATTGCCAACACTTTGTGCAACTCCCGCCAACTCTATGACCAATGATTGAT	1464
Db	1073	GTAAGCTTTGTTCTCATGAACAGTGTATGCAGCCCCCACTACTATGGCTCTGAAATTAAT	1132
QY	1465	GACAGTATGATCTGTGCAGGAATCTTCAGAAACCTGGGCAAGACACCTGCCAGGTGAC	1524
Db	1133	TATAAAATGCTGTGCTGGGACCCAGAGTGGAAAC---AGATTCTGCAAGGCGAT	1189
QY	1525	TCTGGAGGCCCTGACCTGTGAGAGGACGGCACCTACTACGTCTATGGGATAGTGAGC	1584
Db	1190	TCTGGAGGACCGCTTATCTGTAACATCGAAGCCGCCCAACTCTGAGTGGATTGTGAGC	1249
QY	1585	TGGGGCTGAGTGTG-----AGAGAGGCCAGGGGTCTACACCCCAAGTTACCAATTC	1638
Db	1250	TGGGGCGGAGGATGTGAGAGAAACAGCCCGGTCTTACACGAGGTCTCACATTC	1309
QY	1639	CTGAATTGGATCAAAGCCACCATCAAAGTGAAAGTGG	1676
Db	1310	CTGAGCTGGATTCAATCCCAATTGGAGAGAGAAAGG	1347

Search completed: May 25, 2004, 13:22:39
Job time : 537 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model
Run on: May 24, 2004, 09:39:54 ; Search time 54 Seconds
(without alignments)
2930.122 Million cell updates/sec

Title: US-09-912-559-3
Perfect score: 3154
Sequence: 1 MFARMSDLVLLMALVGKT.....TQVTKFLNWKATIKSSGGF 560

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues
Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database : A_Geneseq_29Jan04:*
- 1: Geneseqp1980s:*
 - 2: Geneseqp1990s:*
 - 3: Geneseqp2000s:*
 - 4: Geneseqp2001s:*
 - 5: Geneseqp2002s:*
 - 6: Geneseqp2003as:*
 - 7: Geneseqp2003bs:*
 - 8: Geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3154	100.0	560	3	AAB42484 Human ORF
2	3154	100.0	560	5	AAO17144 Human bio
3	3143	99.7	560	5	AAO17145 Human bio
4	833.5	26.4	655	2	AAR53962 Hepatocyt
5	832.5	26.4	655	2	AAR89197 Human hep
6	812.5	25.8	441	1	AAP70643 Modified
7	812.5	25.8	473	1	AAP70642 Modified
8	807.5	25.6	689	7	ADE07850 Novel pro
9	802	25.4	439	2	AAR09260 t-PA vari
10	798.5	25.3	434	2	AAR13922 Delta (92-
11	797	25.3	439	2	AAR09261 t-PA vari
12	795	25.2	439	2	AAR09258 t-PA vari
13	794	25.2	439	2	AAR09259 t-PA vari
14	793	25.1	623	1	AAP71742 t-PA with
15	792	25.1	623	1	AAP71740 t-PA with
16	786	24.9	527	2	AAR70858 Human t-P
17	786	24.9	562	2	AAR41545 Modified
18	785	24.9	518	2	AAR74684 t-PA mute
19	785	24.9	522	2	AAR13917 Delta (46
20	784	24.9	527	2	AAR70893 Human t-P
21	783	24.8	521	2	AAR74678 t-PA mute
22	782.5	24.8	483	2	AAR09257 t-PA vari
23	782	24.8	522	2	AAR74683 t-PA vari
24	781.5	24.8	483	2	AAR09252 t-PA vari
25	781	24.8	519	2	AAR74687 t-PA mute

26	781	24.8	527	2	AAR70861 Human t-P
27	781	24.8	527	2	AAR70856 Human t-P
28	781	24.8	527	2	AAR70869 Human t-P
29	780.5	24.7	483	2	AAR09246 t-PA vari
30	780	24.7	519	2	AAR74680 t-PA mute
31	780	24.7	522	2	AAR13920 Delta (46
32	780	24.7	527	2	AAR70874 Human t-P
33	780	24.7	527	2	AAR70859 Human t-P
34	779.5	24.7	483	2	AAR70889 Human t-P
35	779.5	24.7	522	2	AAR74685 t-PA mute
36	779	24.7	527	2	AAR70896 Human t-P
37	779	24.7	527	2	AAR70904 Human t-P
38	779	24.7	527	2	AAR70891 Human t-P
39	778.5	24.7	483	2	AAR09250 t-PA vari
40	778.5	24.7	483	2	AAR09251 t-PA vari
41	778.5	24.7	562	2	AAR23802 Zymogen-1
42	778	24.7	518	2	AAR74679 t-PA mute
43	778	24.7	522	2	AAR14486 Delta (46
44	778	24.7	527	2	AAR70876 Human t-P
45	778	24.7	527	2	AAR70865 Human t-P

ALIGNMENTS

RESULT 1
AAB42484
ID: AAB42484 standard; protein; 560 AA.
XX AAB42484;
XX
DT 08-FEB-2001 (first entry)
XX
DE Human ORFX ORF2248 polypeptide sequence SEQ ID NO:4496.
XX
KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW vulnery; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW hypotensive; dermatological; immunosuppressive; antinflammatory;
KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW neurodegenerative disorder; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW cardiovascular disease; systemic lupus erythematosus; infection;
KW cholesterol ester storage; severe combined immunodeficiency; asthma;
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW bone damage; cartilage damage; antinflammatory disease; coagulation;
KW thrombosis; contraceptive.
XX Homo sapiens.
XX WO200058473-A2.
XX
PD 05-OCT-2000.
XX
XX 31-MAR-2000; 2000WO-US008621.
XX
XX 31-MAR-1999; 99US-0127607P.
XX 02-APR-1999; 99US-0127636P.
XX 05-APR-1999; 99US-0127728P.
XX 30-MAR-2000; 2000US-00540763.
XX (CUFA-) CURAGEN CORP.
XX
XX Shinkets RA, Leach M;
XX WPI; 2000-602362/57.
XX N-PSDB; AAC76693.
XX
XX Novel nucleic acids and peptides derived from open reading frame X,
XX useful for treating e.g. cancers, proliferative disorders,
XX neurodegenerative disorders and cardiovascular disease.

XX PS Claim 11; Page 3681-3683; 5507pp; English.

XX CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,

CC which represent the human ORFX open reading frames 1 to 3161. The ORFX

CC sequences have activities such as: cytostatic; hepatotropic; vulnery;

CC antiproliferative; antiparkinsonian; neurotropic; neuroprotective; osteopathic;

CC anticancer; antidiabetic; immunosuppressant; immunostimulant;

CC cardiant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive;

CC dermatological; immunosuppressive; antinflammatory; antibacterial;

CC antiviral; antifungal; antineumatic; antithyroid; and antianemic. The

CC sequences can be used for determining the presence of or predisposition

CC to or preventing or treating pathological conditions associated with an

CC ORFX-associated disorder. The nucleic acids can be used to express ORFX

CC proteins in gene therapy vectors. The proteins and nucleic acids may be

CC used to treat cancers, proliferative disorders, neurodegenerative

CC disorders, osteoarthritis, graft vs host disease, cardiovascular disease,

CC diabetes mellitus, hypertension, hypothyroidism, cholesterol ester

CC storage, systemic lupus erythematosus, severe combined immunodeficiency

CC (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune

CC disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and

CC cartilage damage, nocturnal haemoglobinuria, antinflammatory disease; to

CC enhance coagulation; to inhibit thrombosis; and as a contraceptive

XX SQ Sequence 560 AA;

Query Match 100.0%; Score 3154; DB 3; Length 560;

Best Local Similarity 100.0%; Pred. No. 7.2e-181; Indels 0; Gaps 0;

Matches 560; Conservative 0; Mismatches 0;

QY 1 MFARMSDLHVLMLVKTACGFSLSLLESLEDPDTPDQYDYSYEDYNOENTSSLT 60

DB 1 MFARMSDLHVLMLVKTACGFSLSLLESLEDPDTPDQYDYSYEDYNOENTSSLT 60

QY 61 HAENPDWYTTEDQADPCPNPCEHGGDCLVHGSTFTCSCLAPFSGNKCQVQNTCKDNPC 120

DB 61 HAENPDWYTTEDQADPCPNPCEHGGDCLVHGSTFTCSCLAPFSGNKCQVQNTCKDNPC 120

QY 121 GRGQCLITQSPYYRCVCKPFTGPGSCQVVPVCRPNPCQNGATCSRHKRSKFTACPD 180

DB 121 GRGQCLITQSPYYRCVCKPFTGPGSCQVVPVCRPNPCQNGATCSRHKRSKFTACPD 180

QY 181 QFKGKFCIEGSDDCYVGDGYSYRGKMRNTVNHQACLYNWSHLLLOENYNNFMEDAETHGI 240

DB 181 QFKGKFCIEGSDDCYVGDGYSYRGKMRNTVNHQACLYNWSHLLLOENYNNFMEDAETHGI 240

QY 241 GEHNFCRNPDADEKPCWCFIKVTNDKWKVEYCDVSACSADQVAYPEESPTPESTKLPGFDS 300

DB 241 GEHNFCRNPDADEKPCWCFIKVTNDKWKVEYCDVSACSADQVAYPEESPTPESTKLPGFDS 300

QY 301 CGKTEIAERKTKRIYGGFKSTAGKHPWQASLQSSPLTISMPQGHFCGGALIHPCWYLTA 360

DB 301 CGKTEIAERKTKRIYGGFKSTAGKHPWQASLQSSPLTISMPQGHFCGGALIHPCWYLTA 360

QY 361 AHCTDIKTRHLKVLGDQDLKKEEFHESFVEKIFKYSHYNERDEIPHNDIALLLKLPV 420

DB 361 AHCTDIKTRHLKVLGDQDLKKEEFHESFVEKIFKYSHYNERDEIPHNDIALLLKLPV 420

QY 421 DGHCALESKYVTKVCLPDGSPFSGSECHISGWGTETGKSRQLLDKAVKLIANTLCNSR 480

DB 421 DGHCALESKYVTKVCLPDGSPFSGSECHISGWGTETGKSRQLLDKAVKLIANTLCNSR 480

QY 481 QLYDHMIDSMICAGNLQKPGDTCQDGGGGLTCEKDGTVYVYGVISWGLECKRPGVY 540

DB 481 QLYDHMIDSMICAGNLQKPGDTCQDGGGGLTCEKDGTVYVYGVISWGLECKRPGVY 540

QY 541 TQVTKFLNWKATIKSESGF 560

DB 541 TQVTKFLNWKATIKSESGF 560

RESULT 2

AAO17144

ID AAO17144 standard; protein; 560 AA.

XX AC AAO17144;

XX DT 13-JUN-2002 (first entry)

XX DE Human blood coagulation factor VII activating protease.

XX KW Human; blood coagulation factor VII activating protease; FSAP;

XX KW single-chain plasminogen activator; bleeding disorder; haematological;

XX KW haemostatic.

XX OS Homo sapiens.

XX PN EP1182258-A1.

XX PD 27-FEB-2002.

XX PF 05-JUL-2001; 2001EP-00115691.

XX PR 26-JUL-2000; 2000DE-01036641.

XX PR 10-OCT-2000; 2000DE-01050040.

XX PR 21-OCT-2000; 2000DE-01052319.

XX PR 12-APR-2001; 2001DE-01018706.

XX PA (AVET) AVENTIS BEHRING GMBH.

XX PI Roenisch J, Stoehr H, Feussner A, Lang W, Weimer T, Becker M;

PI Kerlich C, Muth-Naumann G;

XX WIPI; 2002-270939/32.

XX DR N-PSDB; AAL45696.

XX PT New nucleic acid encoding mutant factor 7 activating protease, useful for

PT diagnosis, treatment and prevention of coagulation disorders, also

PT related protein and antibodies.

XX Claim 4; Page 17-19; 27pp; German.

XX The present invention relates to a mutant of the DNA sequence encoding

CC the protease (FSAP) that activates blood coagulation factor VII (FVII)

CC and single-chain plasminogen activator, where at least one of the base

CC changes G to C at nucleotide 1177 and G to A at nucleotide 1601 is

CC present. The mutant sequences can be used in the treatment and prevention

CC of bleeding disorders associated with inherited or acquired defects of

CC blood coagulation factors V, VIII, IX, XI, XII, von Willebrand's

CC factor and/or with antibodies against any of these proteins. The present

CC sequence is the human FSAP protein

XX SQ Sequence 560 AA;

Query Match 100.0%; Score 3154; DB 5; Length 560;

Best Local Similarity 100.0%; Pred. No. 7.2e-181; Indels 0; Gaps 0;

Matches 560; Conservative 0; Mismatches 0;

QY 1 MFARMSDLHVLMLVKTACGFSLSLLESLEDPDTPDQYDYSYEDYNOENTSSLT 60

DB 1 MFARMSDLHVLMLVKTACGFSLSLLESLEDPDTPDQYDYSYEDYNOENTSSLT 60

QY 61 HAENPDWYTTEDQADPCPNPCEHGGDCLVHGSTFTCSCLAPFSGNKCQVQNTCKDNPC 120

DB 61 HAENPDWYTTEDQADPCPNPCEHGGDCLVHGSTFTCSCLAPFSGNKCQVQNTCKDNPC 120

QY 121 GRGQCLITQSPYYRCVCKPFTGPGSCQVVPVCRPNPCQNGATCSRHKRSKFTACPD 180

DB 121 GRGQCLITQSPYYRCVCKPFTGPGSCQVVPVCRPNPCQNGATCSRHKRSKFTACPD 180

QY 181 QFKGKFCIEGSDDCYVGDGYSYRGKMRNTVNHQACLYNWSHLLLOENYNNFMEDAETHGI 240

DB 181 QFKGKFCIEGSDDCYVGDGYSYRGKMRNTVNHQACLYNWSHLLLOENYNNFMEDAETHGI 240

QY 241 GEHNFCRNPDADEKPCWCFIKVTNDKWKVEYCDVSACSADQVAYPEESPTPESTKLPGFDS 300

241 GEHNFNPNDAKEKPCFIKVTNDKVKWEYCDVSACSADQVAYPEESPTSTKLPGFDS 300
301 CGKTEIAERKRIYGGFKSTAGKHPQWASLQSSLPITISMPQGHFCGALIHPCWVLT 360
301 CGKTEIAERKRIYGGFKSTAGKHPQWASLQSSLPITISMPQGHFCGALIHPCWVLT 360
361 AHCTDIKTRHLKVLGDQDLKKEEFHQSFVEKIFKYSYNERDEIPHNDIALLLKPV 420
361 AHCTDIKTRHLKVLGDQDLKKEEFHQSFVEKIFKYSYNERDEIPHNDIALLLKPV 420
421 DGHCALESKYVKTVCPLPGSPSGSECHISGNGVTETGKGRQLLDKVKLIANTLNSR 480
421 DGHCALESKYVKTVCPLPGSPSGSECHISGNGVTETGKGRQLLDKVKLIANTLNSR 480
481 QLYDHMIDDSMTCAGNLQKPGQDTCCQDGGGPLETCEKDGTYVYVGVSWGLECGKRPVY 540
481 QLYDHMIDDSMTCAGNLQKPGQDTCCQDGGGPLETCEKDGTYVYVGVSWGLECGKRPVY 540
541 TQVTKFLNWKATIKSES GF 560
541 TQVTKFLNWKATIKSES GF 560

RESULT 3
AAO17145
ID AAO17145 standard; protein; 560 AA.
XX
AC AAO17145;
XX
DT 13-JUN-2002 (first entry)
XX
DE Human blood coagulation factor VII activating protease mutant.
XX
KW Human; blood coagulation factor VII activating protease; FSAP;
KW single-chain plasminogen activator; bleeding disorder; haematological;
KW haemostatic; mutant; mutein.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN EP118258-A1.
XX
PD 27-FEB-2002.
XX
XX 05-JUL-2001; 2001EP-00115691.
XX
XX 26-JUL-2000; 2000DE-01036641.
XX 10-OCT-2000; 2000DE-01050040.
XX 21-OCT-2000; 2000DE-01052319.
XX 12-APR-2001; 2001DE-01018706.
XX
XX (AVET) AVENTIS BEHRING GMBH.
XX
XX Roemisch J, Stoehr H, Feussner A, Lang W, Weimer T, Becker M;
XX Nerlich C, Muth-Naumann G;
XX
XX WPI; 2002-270939/32.
XX N-PSDB; AAL45697.
XX
XX New nucleic acid encoding mutant factor 7 activating protease, useful for
XX diagnosis, treatment and prevention of coagulation disorders, also
XX related protein and antibodies.
XX
XX Disclosure; Page 20-22; 27pp; German.
XX
XX The present invention relates to a mutant of the DNA sequence encoding
XX the protease (FSAP) that activates blood coagulation factor VII (FVII)
XX and single-chain plasminogen activator, where at least one of the base
XX changes G to C at nucleotide 1177 and G to A at nucleotide 1601 is
XX present. The mutant sequences can be used in the treatment and prevention
XX of bleeding disorders associated with inherited or acquired defects of
XX blood coagulation factors V, VIII, IX, X, XI, XII, von Willebrand's
XX factor and/or with antibodies against any of these proteins. The present

CC sequence is the mutant human FSAP protein
XX
SQ Sequence 560 AA;
Query Match 99.7%; Score 3143; DB 5; Length 560;
Best Local Similarity 99.6%; Pred. No. 3.3e-180;
Matches 569; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 MFARMSDLHLVLLMALVGTACGFSLSLSLLESIDPDWTPDQYDYSYEDYNQEBNTSSTLT 60
DB 1 MFARMSDLHLVLLMALVGTACGFSLSLSLLESIDPDWTPDQYDYSYEDYNQEBNTSSTLT 60
QY 61 HAENPDWYTTEDQADPCQNPCEHGGDCLVHGSDTFTCSCLAPFSGNCKQVQNTCKDNPC 120
DB 61 HAENPDWYTTEDQADPCQNPCEHGGDCLVHGSDTFTCSCLAPFSGNCKQVQNTCKDNPC 120
QY 121 GRQCLITQSPPYRCVCKHPYTGPSCSQVPCVPCRENPQNGATCSRHKRSKFTCACPD 180
DB 121 GRQCLITQSPPYRCVCKHPYTGPSCSQVPCVPCRENPQNGATCSRHKRSKFTCACPD 180
QY 181 QFKGKFCIGSDDCYVGDGYSYRGKKNRTVNOHACLYWNSHLLLOENYNNFMEDAETHGI 240
DB 181 QFKGKFCIGSDDCYVGDGYSYRGKKNRTVNOHACLYWNSHLLLOENYNNFMEDAETHGI 240
QY 241 GEHNFNPNDAKEKPCFIKVTNDKVKWEYCDVSACSADQVAYPEESPTSTKLPGFDS 300
DB 241 GEHNFNPNDAKEKPCFIKVTNDKVKWEYCDVSACSADQVAYPEESPTSTKLPGFDS 300
QY 301 CGKTEIAERKRIYGGFKSTAGKHPQWASLQSSLPITISMPQGHFCGALIHPCWVLT 360
DB 301 CGKTEIAERKRIYGGFKSTAGKHPQWASLQSSLPITISMPQGHFCGALIHPCWVLT 360
QY 361 AHCTDIKTRHLKVLGDQDLKKEEFHQSFVEKIFKYSYNERDEIPHNDIALLLKPV 420
DB 361 AHCTDIKTRHLKVLGDQDLKKEEFHQSFVEKIFKYSYNERDEIPHNDIALLLKPV 420
QY 421 DGHCALESKYVKTVCPLPGSPSGSECHISGNGVTETGKGRQLLDKVKLIANTLNSR 480
DB 421 DGHCALESKYVKTVCPLPGSPSGSECHISGNGVTETGKGRQLLDKVKLIANTLNSR 480
QY 481 QLYDHMIDDSMTCAGNLQKPGQDTCCQDGGGPLETCEKDGTYVYVGVSWGLECGKRPVY 540
DB 481 QLYDHMIDDSMTCAGNLQKPGQDTCCQDGGGPLETCEKDGTYVYVGVSWGLECGKRPVY 540
QY 541 TQVTKFLNWKATIKSES GF 560
DB 541 TQVTKFLNWKATIKSES GF 560

RESULT 4
AAR53962
ID AAR53962 standard; protein; 655 AA.
XX
AC AAR53962;
XX
DT 25-MAR-2003 (revised)
DT 06-JAN-1995 (first entry)
XX
DE Hepatocyte growth factor converting protease.
XX
XX Hepatocyte growth factor; protease; cleavage; active; inactive;
XX precursor.
XX
XX Homo sapiens.
XX
XX EP596524-A2.
XX
XX 11-MAY-1994.
XX
XX 05-NOV-1993; 93EP-00117988.
XX
XX 05-NOV-1992; 92JP-00296133.
XX 20-NOV-1992; 92JP-00312234.
XX

```

PR 20-NOV-1992; 92JP-00312242.
XX (MITU) MITSUBISHI KASEI CORP.
XX
XX Shimomura T, Yamada K, Morimoto Y, Kitamura N, Miyazawa K;
XX
XX WPI: 1994-152921/19.
XX N-PSDB; AAQ63951.
XX
XX Hepatocyte growth factor converting protease and precursor and gene
XX encoding them - for producing active two chain HGF from inactive single
XX chain HGF.
XX
XX Claim 12; Page 21-24; 30pp; English.
XX
XX Hepatocyte growth factor converting protease is capable of converting
XX inactive single chain hepatocyte growth factor (HGF) into active two
XX chain HGF by cleavage at a specific site. (Updated on 25-MAR-2003 to
XX correct FN field.)
XX
XX Sequence 655 AA;
XX
XX Query Match 26.4%; Score 833.5; DB 2; Length 655;
XX Best Local Similarity 33.6%; Pred. No. 6.5e-42;
XX Matches 192; Conservative 89; Mismatches 220; Indels 71; Gaps 21;
XX
XX QY 31 ESLDPDTPDYDYSYEDYNGEENTSTLTHANPDW----YYTEDQA----- 74
XX   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX Db 99 QALTEDGEPCHFPFRYGGRLHACTSGSAHRK---WCATTHNYDRDRAWGYCVETATPPP 155
XX
XX QY 75 -----DPCQNPCEHGGDC--LVHGGTFTCSCLAPFSGNKC--QRVQNTCK-----D 117
XX
XX Db 156 GGPAAALDPCASGPCLINGSCSNTQDQSYHCSPRAFTKDGCTEKCPDETRVEYLEGGD 215
XX
XX QY 118 NFGRGQCLITQSPRYACVCKHYTPGSCQVV--PVCEPBPQCGATCSHKKERSKFT 175
XX
XX Db 216 RWARVRQCHVEQ-----CEC---FGGRTWCEGTRHTACLSSPCLNGGTCHLIVATGTV 266
XX
XX QY 176 CACPDPQFKGFCIEIGDD-CYVGDGYSYRGKMRNTVNOHACLWNHSHLLQENYNMFED 234
XX   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX Db 267 CACPDPGFAGRLCNIEPDERCFLNGTGYRGVASTSAGLSCLAWNSDLLYQELHVDVSVA 326
XX
XX QY 235 AETHGIGEHNFENPDADEKPKCFIKYTNDKVXWECDSACSAA--QDVAYPEESPTSPS 292
XX   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX Db 327 AALLGLGPHAYCRNPDNDRPWCYV-VADSALSWEYCRLEACBSLTRVQLSPDLLLATIPE 385
XX
XX QY 293 TKLPDPSGCKTEIAERKIK-RYVGFKFSTAGKHPQWASLQSSLP.LTISMPQGHFCGGL 351
XX   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX Db 386 PASPGRCACGRRHKKRTFLRPRIIGGSSSLPGSHPLAAIYG-----DSFCAGSL 436
XX
XX QY 352 IHPCXVLTAAHCTDIKTBH--LKVVLGDQDLKKEEFHEQSFREKIFKYSHYNERDEIPH 409
XX   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX Db 437 VTCWVVSAAHCFSSHSPRDSVSVVLQGHFENRTTDTVTFGFEKIPILISVFNPSDH 496
XX
XX QY 410 NDIALKLKVPDGHCALESKYKTVCLPD--GSFPGSGECHISGWG-VTETGKG-SROLL 465
XX   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX Db 497 -DIVLIRLKKKDRCATRSQFVQPICLPEPGSTFPAGHKCQIAGWGHLDENVSGYSSILR 555
XX
XX QY 466 DAKVKLIANTLGNRSQLYDHDIMDSMTCAGNLKPQGDTCQGGSGGPLTCEKDGTVYVYG 525
XX   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX Db 556 EALVPLVADHKSSFEVIGALISPNMLCAGVFDC-SDACQGSQGLACERKNVAYLYG 614
XX
XX QY 526 IVSWGLECKG--RPGVYQVTKFLNLWKATIK 555
XX   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX Db 615 IISWGGCGRLHKPGVYTRVANYVDINDIR 646
XX
XX RESULT 5
XX AAR89197
XX ID AAR89197 standard; protein; 655 AA.
XX
XX AAR89197;
XX

```

```
QY 404 RDEIPHNDIALKLPVDGHCALSKYKVTCLPD--GSFPSSGEGCHISGNG-VTETGK 460
Db 491 FNPSPH-DLVLRILKKGDRCATRSQFVQPCLPBPSTFPAGHKCQIAGHGLDENVSG 549
QY 461 -SRQLLDKAVKLIANTLNSRQLYDHMIDDSMICAGNLQKPGQDTCCQSDSGPLTCEKDG 519
Db 550 YSSSLREALPLVADHKCSSPEVYGADISPNMLCAGYFDCK-SDACQSDSGPLACEKNG 608
QY 520 TYYVYGIYSWGLECGK--RPGVYTVTKFLNWKATIK 555
Db 609 VAYLYGISWGDGGRLLKPGVYTRVANYVDWINDRIIR 646

RESULT 6
AAP70643
ID AAP70643 standard; protein; 441 AA.
XX
AC AAP70643;
XX
DT 25-MAR-2003 (revised)
DT 10-APR-1991 (first entry)
XX
DE Modified tissue plasminogen activator.
XX
KW Tissue plasminogen activator; kringle domain; embolism; thrombosis;
KW stroke.
XX
OS Homo sapiens.
XX
PN AU8661804-A.
XX
PD 05-MAR-1987.
XX
PF 25-AUG-1986; 86AU-00061804.
XX
PR 26-AUG-1985; 85US-00769298.
XX
PA (ELIL ) LILLY & CO ELI.
XX
PI Bang NU, Little SP, Schonher BE, Weigel BJ;
XX
DR WPI; 1987-108842/16.
XX
DR N-PSDB; AAN70992.
XX
PT New modified tissue plasminogen activator - with new encoding DNA, new
PT DNA expression vector, useful for treating vascular disorders, e.g.,
PT pulmonary embolism arterial thrombosis.
XX
PS Claim 9; Page 122; 164pp; English.
XX
CC The modified t-PA has all/part of the kringle domains of native t-PA
CC removed. The t-PA has functional properties superior to those of native t
CC -PA. It retains fibrin binding properties and interacts more slowly and
CC inefficiently with plasminogen activator inhibitor(s) compared to native
CC t-PA. It is obtd. in large amts. from a prokaryotic host. Modified t-PA
CC is used for treating vascular disorders, eg deep vein thrombosis,
CC pulmonary embolism, peripheral arterial thrombosis, disseminated
CC intravascular coagulation, emboli from the heart or peripheral arteries,
CC acute myocardial infarction, thrombotic strokes or fibrin deposits
CC associated with invasive cancers. t-PA is used at a dosage of 250000 to
CC 500000 units at a loading dose or in a deep vein thrombosis-pulmonary
CC embolism, or 250000-750000 units over 30-90 mins. in acute myocardial
CC infarction. (Updated on 25-MAR-2003 to correct PA field.)
XX
SQ Sequence 441 AA;

Query Match 25.8%; Score 812.5; DB 1; Length 441;
Best Local Similarity 37.7%; Pred. NO. 8e-41;
Matches 177; Conservative 70; Mismatches 163; Indels 59; Gaps 14;

QY 112 QNTCKNPGCRGQCLITQSPPYR-----CVCKHPYTGPSQVVPV--CRPNPC 159
Db 5 QVICRDE--XTQVIYQHQSWLRPVLRNSRVEYWCN---SGRACHSVPKSCSEPRC 58
```

```
QY 160 QNGATCSRHKRSKFTACPDQKGFCEIGSDDCYVGDGYSGYRGKQNRFTVQNHACLYWN 219
Db 59 FNGGTQOALYFSDFYCQCPGEGFAGKSCIEGNSDCYFGNGSAVGRTHSLTESGASCLPN 118
QY 220 SHLLGENTNMFMEDAEHIGIEHNPCRNPDADERKFWCFIKVTNDKVKWEYCDVSACSAQ 279
Db 119 SMILIGKVTAQNPASAALGLGKHNPCRNPDGDGAKFWCHV-LKNRRLTWYCDVPSCS-- 175
QY 280 DVAYPEESPTEPSTKLPGDFSCGKTEIAERKIKRIYGGFKSTAGKHPWQASLQSSLP 339
Db 176 -----TCGLRYSQPF-RIKGGLFADIASHPWQAIFAK--HRR 212
QY 340 SMPQGHFGGALHPQVLTAAHCTD--IKTRHLKVLGDQDLKKEEFHQSFVEKIKFK 397
Db 213 SPGERFLCGGILISSCWILSAAHCFQERFPFPHLTVLIGRTYRVWPEBEQFEVSKYIV 272
QY 398 YSHYNERDEIPHNDIALKLPVDGHCALSKYKVTCLP--DGSPSSGEGCHISGNGVT 455
Db 273 HKEFD--DDTYDNDIALQLKSDSSRCAQESSVVRVCLPPADLQLPDWTECELSGYKH 330
QY 456 ETGKG--SRQLLDKAVKLIANTLNSRQLYDHMIDDSMICAGNLQKPG-----QDTCQGD 508
Db 331 EALSPFYSERLKEAHVRLYPSRSCTSQHLINRTVTDNMLCAGDTRSGGPQANLHDACQGD 390
QY 509 SGGPLCEKDGTYVYGIYSWGLECGK--RPGVYTVTKFLNWKATIK 555
Db 391 SGGPLVCLNDGRMTLVYGISWGLECGCKQKQVPGVYTKVTNYLDWIRDNR 439

RESULT 7
AAP70642
ID AAP70642 standard; protein; 473 AA.
XX
AC AAP70642;
XX
DT 25-MAR-2003 (revised)
DT 10-APR-1991 (first entry)
XX
DE Modified tissue plasminogen activator.
XX
KW Tissue plasminogen activator; kringle domain; embolism; thrombosis;
KW stroke.
XX
OS Homo sapiens.
XX
PN AU8661804-A.
XX
PD 05-MAR-1987.
XX
PF 25-AUG-1986; 86AU-00061804.
XX
PR 26-AUG-1985; 85US-00769298.
XX
PA (ELIL ) LILLY & CO ELI.
XX
PI Bang NU, Little SP, Schonher BE, Weigel BJ;
XX
DR WPI; 1987-108842/16.
XX
DR N-PSDB; AAN70991.
XX
PT New modified tissue plasminogen activator - with new encoding DNA, new
PT DNA expression vector, useful for treating vascular disorders, e.g.,
PT pulmonary embolism arterial thrombosis.
XX
PS Claim 9; Page 122; 164pp; English.
XX
CC The modified t-PA has all/part of the kringle domains of native t-PA
CC removed. The t-PA has functional properties superior to those of native t
CC -PA. It retains fibrin binding properties and interacts more slowly and
CC inefficiently with plasminogen activator inhibitor(s) compared to native
CC t-PA. Obtd. in large amts. from a prokaryotic host. Modified t-PA is used
CC for treating vascular disorders, eg deep vein thrombosis, pulmonary
```

CC embolism, peripheral arterial thrombosis, disseminated intravascular
 CC coagulation, emboli from the heart or peripheral arteries, acute
 CC myocardial infarction, thrombotic strokes or fibrin deposits associated
 CC with invasive cancers. t-PA is used at a dosage of 250000 to 500000
 CC units at a loading dose or in a deep vein thrombosis-pulmonary embolism,
 CC or 250000-750000 units over 30-90 mins. in acute myocardial infarction.
 CC (Updated on 25-MAR-2003 to correct PA field.)
 XX
 SQ Sequence 473 AA;

Query Match 25.8%; Score 812.5; DB 1; Length 473;
 Best Local Similarity 37.7%; Pred. No. 8.5e-41;
 Matches 177; Conservative 70; Mismatches 163; Indels 59; Gaps 14;
 QY 112 QNTCKDNPCGRGQCLITQSPYYR-----CVCKHPYTGSCSOWVPV--CRNPNC 159
 DB 38 QVICRDE---KTMQVYQHQSLRPVLRNREVEICWCVN---SGRAQCHSVFVSCSEPRC 91
 QY 160 QNGATCSRHRKRSKFTACPDQFKGFCIEGSDCYVDGYSYRGKNNRTVNQHACLYWN 219
 DB 92 FNGGTCQALYFSDVQCQPEFAGKSCIEGNSDCYFGNGSAYRGTHTSLTESGASCLPWN 151
 QY 220 SHLLQENYMWEDAEATHGIEHNFERNPDADKPCWCFIKVTNDKVKWEYCDVSAQ 279
 DB 152 SMILIGKVTAQNPQAQALGLGKHNKCRNPDGDAKPCHV-LKNERLTWYCDVPSCS-- 208
 QY 280 DVNYPEESTREPKLPGDSCGKTEIAERKRIYGGPKSTAGKHPWQASLQSLPLTI 339
 DB 209 -----TCGLRYSQPOF-RIKGLFADIASHPWQAIFAK--HRR 245
 QY 340 SMPQGHFCGGLIHPQWLTAAHCTD--IKTRHLKVLGVDQDLKKEEFHEOSRVEKIFK 397
 DB 246 SPGERFLCGILLISSCWLSAAHCFQERFPPHLLTVILGRYRVVPGEEKQKEVEKYIV 305
 QY 398 YSHYNERDEIPNDIALLKKEVDGHCALSKYKVTCLP--DQSPFGSECHISGWVIT 455
 DB 306 HKEFD--DDTYNDIALLQLKSDSRCAQESSVVRTVCLPPADILQLPDWTCELSGYGKH 363
 QY 456 ETKGK--SRQLLDKVKLIANTLNSRQLYDHMDSDSMICAGNLQKPG-----QDTCCGD 508
 DB 364 EALSPYSERLKEAHVRLVPSRCTSOHLNRTVTDNMLCAGDTRSGPQANLHDAQQD 423
 QY 509 SGGLTCEKDGYYVYGVISWGLECGKR--FGVYTVTKFLNWKATIK 555
 DB 424 SGGLVCLNDGRNLTGLVGIISWGLGCGQKDVGVYTKVTNLDLWDNR 472

RESULT 8
 ADE07850
 ID ADE07850 standard; protein; 689 AA.
 XX
 AC ADE07850;
 XX
 DT 29-JAN-2004 (first entry)
 XX
 DE Novel protein (useful for identifying genetic disorders) #5.
 KW novel gene; novel protein; tissue marker; molecular weight marker;
 KW chromosome marker; genetic disorder.
 XX Unidentified.
 XX
 XX WO2003054152-A2.
 XX
 XX 03-JUL-2003.
 XX
 XX 10-DEC-2002; 2002WQ-US039555.
 XX
 XX 10-DEC-2001; 2001US-0339739P.
 PR 11-DEC-2001; 2001US-0339453P.
 PR 14-MAR-2002; 2002US-0365091P.
 PR 14-MAR-2002; 2002US-0365384P.
 PR 12-APR-2002; 2002US-0372381P.

PR 12-APR-2002; 2002US-0372615P.
 PR 22-APR-2002; 2002US-00128558.
 PR 24-APR-2002; 2002US-0376045P.
 XX (HYSE-) HYSEQ INC.
 XX
 XX Tang YT, Asundi V, Goodrich RM, Ren F, Zhang J, Zhao QA, Wang J;
 PI Ghosh M, Xue AJ, Wehrman T, Weng G, Zhou P, Drmanac RT, Wang Z;
 PI Ma Y, Wang B, Chen R, Xu C, Boyle BJ;
 XX
 XX WPI; 2003-569235/53.
 DR N-PSDB; ADE06939.
 XX
 PT New polynucleotides, useful for expressing recombinant proteins for
 PT analysis, characterization or therapeutic use, or as markers for tissues
 PT in which the corresponding protein is preferentially expressed.
 XX
 XX Claim 20; SEQ ID NO 916; 1177pp; English.
 XX
 CC The invention comprises the amino acid and coding sequences of novel
 CC proteins. The DNA and protein sequences of the invention are useful as:
 CC markers for tissues in which the corresponding protein is preferentially
 CC expressed; as molecular weight markers on gels; as chromosome markers or
 CC tags; to identify chromosomes or to map related gene positions; and to
 CC compare with endogenous DNA sequences in patients to identify potential
 CC genetic disorders. The present amino acid sequence represents a protein
 CC of the invention.
 XX
 SQ Sequence 689 AA;
 Query Match 25.6%; Score 807.5; DB 7; Length 689;
 Best Local Similarity 31.3%; Pred. No. 2.4e-40;
 Matches 192; Conservative 89; Mismatches 213; Indels 119; Gaps 23;
 QY 31 ESLPDMTPDDYYSYDYNQEBNTSTLTHANPW-----YTEDQA----- 74
 DB 99 QALTDEGRPCFPFRYGRMLHACTSEGSAHRK---WCATHNYDRDRAGYCVETATPPP 155
 QY 75 -----DPCQNPCEHGGDC--LVHGSTFTCSCLAPFGNKKQKQVCTCKDNPCGRGQCL 126
 DB 156 GCPAALDFCAGSPCLNGSGCSNTQDPQSVHCSCPRAFTG-----KD--CGTEKCF 203
 QY 127 ITQSPFY-----RCVCKHPTTGPSCSVV--PVRPNPCNGATCSRHK 169
 DB 204 DETRYEYLEGDRWARVRQGHVEQEC---FGRTWCETRTHTACLSPPCLNGTCHLIV 260
 QY 170 RRSKFTACPDQFKGFCIEGSD--CYVDGYSYRGKNNRTVNQHACLYWNHLLQENY 228
 DB 261 ATGTTVCACPPGAFAGRLCNIEPDERCFLNGTGYRGVASTASGLSCLAWNSDLLYQELH 320
 QY 229 NMFMEDAETHGIEHNFERNPDADKPCWCFIKVTNDKVKWEYCDVSAQ----- 276
 DB 321 VDSVGAALLGLGPHAYCRNPDNDERPCYV--VDSALSWEYCRLEACGPRRQPGALMW 379
 QY 277 -----SAQDVAYPEESP-----TEPSTKLPG-----FDSCKGTEIAER 309
 DB 380 GSLTVLGLAVGFAAGATKASVCPALALTEPRDRASNPILQIWKLRACGRHKRT 439
 QY 310 KIK-RYVGGFKTAGKHPQASLOSSPLTISMPQGHFCGGLIHPQWLTAAHCTDIKT 368
 DB 440 FLRPLIIGSSSLFGSHPLAAIYIG-----DSFCAGSLVHTCVVWSAAHCFSHRD 490
 QY 369 RHLKYLGDQDLKKEEFHEOSRVEKIFKYSHYNERDEIPNDIALLKKEVDGHCALSK 428
 DB 491 -SVSVVLGQHFFNRITDVTQTFGIEKYIPYTLKSVFNPSDH-DLVLTIRLKKGDRCATRS 548
 QY 429 KYVKTVCCLPD--GSFPBSGSECHISGCV-VTETGK-SRQLLDKVKLIANTLNSRQLYD 484
 DB 549 QFVQPICLPFGSTFPAGHKCQIAGWGHLDENVSGYSSSLREALVPLVDHKKCSSEVYG 608
 QY 485 HMIDDSMT CAGNLQKPGDTCQSDSGGFLTCEKDGTYVYGVISWGLECGK--RPGVYVQ 542
 DB 609 ADISPNMLCAGYFDCX-SDACQDGGGFLACENGVAVLYGIISWGLGCGQRLHKPGVYTR 667

QY 543 VTKFLNWKATIK 555
 Db 668 VANYVDWINDQIR 680

RESULT 9
 AAR09260
 ID AAR09260 standard; protein; 439 AA.
 XX
 AC AAR09260;
 DT 24-FEB-1993 (first entry)
 DE t-PA variant d92-179, N184D, I210R, G211A, K212R, V213R, T252R, F305H.
 XX
 KW Tissue plasminogen activator; zymogen; clot; plasma; plasmin.
 XX
 OS Homo sapiens.
 FH Key
 FT Misc-difference 91..92 Location/Qualifiers
 FT /note= "amino acids 92-179 of the wild-type, mature t-PA
 FT have been deleted"
 FT Misc-difference 96 /label= substitution
 FT /note= "N184 of the wild-type, mature t-PA has been
 FT substituted for D"
 FT Misc-difference 122 /label= substitution
 FT /note= "I210 of the wild-type, mature t-PA has been
 FT substituted for R"
 FT Misc-difference 123 /label= substitution
 FT /note= "G211 of the wild-type, mature t-PA has been
 FT substituted for A"
 FT Misc-difference 124 /label= substitution
 FT /note= "K212 of the wild-type, mature t-PA has been
 FT substituted for R"
 FT Misc-difference 125 /label= substitution
 FT /note= "V213 of the wild-type, mature t-PA has been
 FT substituted for R"
 FT Misc-difference 164 /label= substitution
 FT /note= "T252 of the wild-type, mature t-PA has been
 FT substituted for R"
 FT Misc-difference 217 /label= substitution
 FT /note= "F305 of the wild-type, mature t-PA has been
 FT substituted for H"
 FT
 PN WO9002798-A.
 XX
 PD 22-MAR-1990.
 XX
 PF 02-SEP-1988; 88US-00240856.
 XX
 PR 02-SEP-1988; 88US-00240856.
 PR 24-JUL-1989; 89US-00384608.
 XX
 PA (GETH) GENENTECH INC.
 XX
 PI Anderson S, Bennett WE, Botstein D, Higgins DL, Paoni NF;
 PI Zoller M;
 XX
 DR WPI; 1990-115987/15.
 XX
 XX Modified tissue plasminogen activator - activated only when proximate to
 PT plasmin at site of clot and not systemically.
 PT
 XX Claim 17; Page 44; 63pp; English.
 PS

XX The t-PA variants represented in AAR09215-63 (and the T252R or N184S
 CC analogues thereof or combinations thereof) and AAR09266-83 have their
 CC fibrinolytic activity turned off when generally in the plasma and
 CC activated when proximate to plasmin at the site of the clot thus
 CC providing specific localised clot therapy
 XX
 SQ Sequence 439 AA;
 Query Match 25.4%; Score 802; DB 2; Length 439;
 Best Local Similarity 37.7%; Pred No. 3.4e-40;
 Matches 177; Conservative 68; Mismatches 165; Indels 60; Gaps 15;
 QY 112 QNTCKDNPCGRQCCLITQSPYYR-----CVCXHPYTPGSCSQVVPV--CRPNPC 159
 Db 3 QVICRDE---KTQMIYQQHQSRLRSLRSNRVEYWCN---SGRAQCHSVVPKCSSEPRC 56
 QY 160 QNGATCSRHKRSKFTCACPDQFKGKFCBEGS-DDCYVDGDYSGYRGKMRVTNQCACLYW 218
 Db 57 FNGGTCQALYFSDFVCCQCPGFAKCEIDTRATCYFGDSAYRGTHSLTESGASCLPW 116
 QY 219 NSHLLAQENYNNFMEDAETHGIGENFNCPNPDADSKPWCIFKVTNDKVKWEYCDVSACSA 278
 Db 117 NSMLRARRYTAQNPSAQAALGLGKKNYCRNPDGDAKPWCHV-LKNRRLRWEYCDVPSCS- 174
 QY 279 QDVAYPEESPTPESTKLPQFDSCKGTEIAERIKRIYGGFKSTAGKHPQWQASLQSLPLT 338
 Db 175 -----TCGLRQYSQPF-RIKGSLFADIASHPQAAIFAK--HR 210
 QY 339 ISMPQGHFCGGALIHPCWVLTAAHCTD--IKTRHLKVLGDQDLKKEEFHESGFRVEKIF 396
 Db 211 RSPGHERLGGILISSCWILSAAHCFQERPPPHLLTVILGRTYRVVPGSEEQKFEVEKYI 270
 QY 397 KYSHYNERDEIPHNDIALLKLPVDGHCALSKYKVTVCPL--DGSFSPSGSECHISGMGV 454
 Db 271 VHKEFD--DQTYDNDIALQLKSDSSRCAQESSVVRVCLPPADLQLPDWTCELSGYK 328
 QY 455 TETGKG--SRQLLDKVKLIANTLCNRSQLYDHMDSDSMICAGNLQKPG-----QDTCOG 507
 Db 329 HEALSPFYSERLKEAHVRLYPSRCTSQHLLNRTVTDNMLCAGDTRSGGPGQANLHDACOG 388
 QY 508 DSGGPLTCEKDGTYVYVIGVSWGLECGKR--PGVYTVQTKFLNWKATIK 555
 Db 389 DSGGPLVCLNDRMTLVGIISWGLGGQKDVPGVYTKVTNYLDWIRDNMR 438

RESULT 10
 AAR13922
 ID AAR13922 standard; protein; 434 AA.
 XX
 AC AAR13922;
 DT 09-JAN-2003 (revised)
 DT 27-NOV-1991 (first entry)
 DE Delta(92-179) and delta(466-470) tPA variant.
 XX
 KW tissue plasminogen activator; infarction; coagulation.
 XX
 OS Synthetic.
 FH Key
 FT Domain Location/Qualifiers
 FT 1..39 /label= Finger_domain
 FT 92..173 /label= Kringle_2
 FT
 XX WO9113149-A.
 XX
 PD 05-SEP-1991.
 XX
 PF 01-MAR-1990; 90US-00486657.
 XX

PR 01-MAR-1990; 90US-00486657.
XX (GETH) GENENTECH INC.
XX Gill JF, Presta LG, Zoller ML;
XX WPI; 1991-281468/38.
DR
XX New tissue plasminogen activator variant used for vascular conditions -
PT e.g. to prevent fibrin deposition of adhesion formation of reformation,
PT deep vein thrombosis, peripheral vascular disease, embolism etc.
XX
PS Claim 12; Page 27; 33pp; English.
XX
CC This tPA variant has a deletion of amino acids Pro(466) to Leu(470) of
CC the corresponding wild-type. The deletion is in the serine protease
CC domain, making the variant more fibrin specific than the wild-type tPA.
CC The variant also has a deletion of amino acids Cys(92) to Asp(179) of the
CC wild-type sequence (i.e. Kingle 1 domain). See also AAR13917-R13921 and
CC AAR14486. (Updated on 09-JAN-2003 to add missing OS field.)
XX
SQ Sequence 434 AA;

Query Match 25.3%; Score 798.5; DB 2; Length 434;
Best Local Similarity 37.6%; Pred. No. 5.4e-40;
Matches 175; Conservative 70; Mismatches 165; Indels 55; Gaps 14;

QY 112 QNTCKDNPCGRGQCLITQSPPPYR-----CVCKHPYTGSCSQQVPPV--CRENPC 159
DB 3 QVICRDE---KTQMIYQCHOSWLRPVLSNRVEYWCN---SGRAQCHSVPVKSCSEPRC 56

QY 160 QNGATCSRHRKSKFTACPDQFKGKFCIEGS-DDCYVGDGYSYRGKQNRVTNQHACLYW 218
DB 57 FNGGTCQALYFSDFCVQCPEGFAGKCEIDTRATCYFGNGSAVYRGTHSLTESGASCLPW 116

QY 219 NSHLLLOENYMFEMDAETHGIGENFCRNPDADEKPCFIKVTNDKVMYCDVSACSA 278
DB 117 NSMILICKYTAQNSAQAALGLGKYNCEPDDGDAKPNCHV-LKMRLETWYCDVPSCS- 174

QY 279 QDVAYPEESFTPESTKLPQFDCSGKTEAERKIRYGGFKSTAGKHPQWASLSPLT 338
DB 175 -----TCGLRQYSQPOF-RIKGLFADIASHFPWQAAIFAK--HR 210

QY 339 ISMPOGHFCGGALIHPCWVLTAAHCTD--IKTRHLKVLVGLDQDKKEEFESQSRVEKIF 396
DB 211 RSPGERFLCGGILISSCWLSAAHCFQERFPPHLLTVILGRYVRVPGEEQKEVEKYI 270

QY 397 KYSHYNERDEIPHNDIALKLPVDGHCALSKYKVTCLP--DGSFPGSECHISGNGV 454
DB 271 VKXEPD--DDTYDNDIALQLKSDSSRCAQESSVVRTVCLPPADLQLPDMTECELSGYK 328

QY 455 TETGKG--SRQLIDAKVKLIANTLNSRQLYDHMIDDSMICAGNLKPGQDTCCQDGGP 512
DB 329 HEALSPFYSERLKEAHVRLYPSRCTSQHLLNRTVTDNMLCAGDTRSGHDAQCGDGGP 388

QY 513 LTCCKDGTYYVGVISWGLECGKR--PGYITQVTKFLMWIKATIK 555
DB 389 LVCNDGRMTLVGIISWGLGCGQKQDVPGYTKVTNYLDWIRDNR 433

RESULT 11
AAR09261
ID AAR09261 standard; protein; 439 AA.
XX
AC AAR09261;
XX
DT 24-FEB-1993 (first entry)
XX
DE t-PA variant d92-179, I210R, G211A, K212R, V213R, T252R, F305H.
XX
KW Tissue plasminogen activator; zymogen; clot; plasma; plasmin.
XX
OS Homo sapiens.

XX Key Location/Qualifiers
FH Misc-difference 91..92
FT /note= "amino acids 92-179 of the wild-type, mature t-PA
FT have been deleted"
FT Misc-difference 122
FT /label= substitution
FT /note= "I210 of the wild-type, mature t-PA has been
FT substituted for R"
FT Misc-difference 123
FT /label= substitution
FT /note= "G211 of the wild-type, mature t-PA has been
FT substituted for A"
FT Misc-difference 124
FT /label= substitution
FT /note= "K212 of the wild-type, mature t-PA has been
FT substituted for R"
FT Misc-difference 125
FT /label= substitution
FT /note= "V213 of the wild-type, mature t-PA has been
FT substituted for R"
FT Misc-difference 164
FT /label= substitution
FT /note= "T252 of the wild-type, mature t-PA has been
FT substituted for R"
FT Misc-difference 217
FT /label= substitution
FT /note= "F305 of the wild-type, mature t-PA has been
FT substituted for H"
XX
PN WO9002798-A.
XX
PD 22-MAR-1990.
XX
PF 02-SEP-1988; 88US-00240856.
XX
PR 02-SEP-1988; 88US-00240856.
PR 24-JUL-1989; 89US-00384608.
XX
PA (GETH) GENENTECH INC.
XX
PI Anderson S, Bennett WE, Botstein D, Higgins DL, Paoni NF;
PI Zoller M;
XX
WPI; 1990-115987/15.
XX
PT Modified tissue plasminogen activator - activated only when proximate to
PT plasmin at site of clot and not systemically.
XX
PS Claim 17; Page 44; 63pp; English.
XX
CC The t-PA variants represented in AAR09215-63 (and the T252R or N184S
CC analogues thereof or combinations thereof) and AAR09266-83 have their
CC fibrinolytic activity turned off when generally in the plasma and
CC activated when proximate to plasmin at the site of the clot thus
CC providing specific localised clot therapy
XX
SQ Sequence 439 AA;

Query Match 25.3%; Score 797; DB 2; Length 439;
Best Local Similarity 37.4%; Pred. No. 6.7e-40;
Matches 176; Conservative 69; Mismatches 165; Indels 60; Gaps 15;

QY 112 QNTCKDNPCGRGQCLITQSPPPYR-----CVCKHPYTGSCSQQVPPV--CRENPC 159
DB 3 QVICRDE---KTQMIYQCHOSWLRPVLSNRVEYWCN---SGRAQCHSVPVKSCSEPRC 56

QY 160 QNGATCSRHRKSKFTACPDQFKGKFCIEGS-DDCYVGDGYSYRGKQNRVTNQHACLYW 218
DB 57 FNGGTCQALYFSDFCVQCPEGFAGKCEIDTRATCYFGNGSAVYRGTHSLTESGASCLPW 116

QY 219 NSHLLLOENYMFEMDAETHGIGENFCRNPDADEKPCFIKVTNDKVMYCDVSACSA 278
DB 117 NSMILICKYTAQNSAQAALGLGKYNCEPDDGDAKPNCHV-LKMRLETWYCDVPSCS- 174

Query Match 25.3%; Score 797; DB 2; Length 439;
Best Local Similarity 37.4%; Pred. No. 6.7e-40;
Matches 176; Conservative 69; Mismatches 165; Indels 60; Gaps 15;

Db 117 NSMILPAREYTAQNPSAQAALGLGKXNYCENPDGDAKPKWCHV-LKNRRLRWECYDVPSCS- 174

QY 279 QDVAYPEESPTBTKLPDFDSCGTEIAERKIKIYGGFKSTAGKHQWQASLOSSPLT 338

Db 175 -----TCGLRQYSQPF-RINGGLFADIASHPWQAIAFAK--HR 210

QY 339 ISMPOGHFCGGALIHPCWVLTAAHCTD--IKTRHLKVLGDQDLKKEEFHQSRVEKIF 396

Db 211 RSPGERHLCGGILISSCWILSAHCFQERPPPHLTVILGRYRVVPGEEQKFEVEKYI 270

QY 397 KYSHYNERDEIPHNIDIALKLPVDGHCALESKYVKTVCPL--DGSFPGSGECHISGMV 454

Db 271 VKHEFD--DDTYNDNIALQLKSDSRCAQESSVVRTVCLPADLQLPDWTCELSGYGK 328

QY 455 TETGKG--SRQLLDAKVLIANTLNSRQLYDHMDDSMICAGNLQKPG-----QDTCCG 507

Db 329 HEALSPFYSERLKEAHVRYPSCRTSQHLLNRTVTDNMLCAGDTRSGGPQANLHDACQ 388

QY 508 DSGGPLTCEKDGTYYVYGVISWGLECGKR--PGVYTQVTKFLNWKATIK 555

Db 389 DSGGPLVCLNDGRMTLVGIISWGLGCGQKDPGVYTKVTNYLDWIRDNR 438

RESULT 12

AAR09258

ID AAR09258 standard; protein; 439 AA.

AC AAR09258;

DT 24-FEB-1993 (first entry)

XX t-PA variant d92-179, F305H.

XX Tissue plasminogen activator; zymogen; clot; plasma; plasmin.

XX Homo sapiens.

XX Key Location/Qualifiers

FT Misc-difference 91..92

FT /notes= "amino acids 92-179 of the wild-type, mature t-PA have been deleted"

FT Misc-difference 217

FT /label= substitution

FT /notes= "F305 of the wild-type, mature t-PA has been substituted for H"

XX WO9002798-A.

XX 22-MAR-1990.

XX 02-SEP-1988; 88US-00240856.

XX 02-SEP-1988; 88US-00240856.

XX 24-JUL-1989; 89US-00384608.

XX (GETH) GENENTECH INC.

XX Anderson S, Bennett WE, Botstein D, Higgins DL, Paoni NF; Zoller M;

XX WPI; 1990-115987/15.

XX Modified tissue plasminogen activator - activated only when proximate to plasmin at site of clot and not systemically.

XX Claim 17; Page 44; 63pp; English.

XX The t-PA variants represented in AAR09215-63 (and the T252R or N184S analogues thereof or combinations thereof) and AAR09266-83 have their fibrinolytic activity turned off when generally in the plasma and activated when proximate to plasmin at the site of the clot thus providing specific localised clot therapy

SQ Sequence 439 AA;

Query Match 25.2%; Score 795; DB 2; Length 439;

Best local Similarity 37.4%; Pred. NO. 8.8e-40;

Matches 176; Conservative 70; Mismatches 164; Indels 60; Gaps 15;

QY 112 QNTCKNPNCRGQCLITQSPPPYR-----CVCKHPYTGSPSCQVVPV--CRPNPC 159

Db 3 QVICRDE--KTQWIVQHQSWLRPVLRSNRVYWCN--SGRAQCHSVPVKSCSEPRC 56

QY 160 QNGATCSRHRKRKFTACPDQKGFCEIGS--DDCYVGDGYSGYRGMNRTVNHACLYW 218

Db 57 FNGGTCQQAALYFSDVQCQPEGFAGKCEIDTRATCYFNGSAYRGTHTSLTESGASCLPW 116

QY 219 NSHLLQENYMFEDAEATHGIGEHNFRCNPDPADKPCWCFIKVTNDKVKWEYCDVSACSA 278

Db 117 NSMILIGKVTYTAQNPASAQALGLGKXNYCENPDGDAKPKWCHV-LKNRRLRWECYDVPSCS- 174

QY 279 QDVAYPEESPTBTKLPDFDSCGTEIAERKIKIYGGFKSTAGKHQWQASLOSSPLT 338

Db 175 -----TCGLRQYSQPF-RINGGLFADIASHPWQAIAFAK--HR 210

QY 339 ISMPOGHFCGGALIHPCWVLTAAHCTD--IKTRHLKVLGDQDLKKEEFHQSRVEKIF 396

Db 211 RSPGERHLCGGILISSCWILSAHCFQERPPPHLTVILGRYRVVPGEEQKFEVEKYI 270

QY 397 KYSHYNERDEIPHNIDIALKLPVDGHCALESKYVKTVCPL--DGSFPGSGECHISGMV 454

Db 271 VKHEFD--DDTYNDNIALQLKSDSRCAQESSVVRTVCLPADLQLPDWTCELSGYGK 328

QY 455 TETGKG--SRQLLDAKVLIANTLNSRQLYDHMDDSMICAGNLQKPG-----QDTCCG 507

Db 329 HEALSPFYSERLKEAHVRYPSCRTSQHLLNRTVTDNMLCAGDTRSGGPQANLHDACQ 388

QY 508 DSGGPLTCEKDGTYYVYGVISWGLECGKR--PGVYTQVTKFLNWKATIK 555

Db 389 DSGGPLVCLNDGRMTLVGIISWGLGCGQKDPGVYTKVTNYLDWIRDNR 438

RESULT 13

AAR09259

ID AAR09259 standard; protein; 439 AA.

AC AAR09259;

DT 24-FEB-1993 (first entry)

XX t-PA variant d92-179, I210R, G211A, K212R, V213R, F305H.

XX Tissue plasminogen activator; zymogen; clot; plasma; plasmin.

XX Homo sapiens.

XX Key Location/Qualifiers

FT Misc-difference 91..92

FT /notes= "amino acids 92-179 of the wild-type, mature t-PA have been deleted"

FT Misc-difference 122

FT /label= substitution

FT /notes= "I210 of the wild-type, mature t-PA has been substituted for R"

FT Misc-difference 123

FT /label= substitution

FT /notes= "G211 of the wild-type, mature t-PA has been substituted for A"

FT Misc-difference 124

FT /label= substitution

FT /notes= "K212 of the wild-type, mature t-PA has been substituted for R"

FT Misc-difference 125

FT /label= substitution

FT /notes= "V213 of the wild-type, mature t-PA has been substituted for R"

```

FT Misc-difference 217
FT /label= substitution
FT /note= "F305 of the wild-type, mature t-PA has been
FT substituted for H"
XX
XX WO9002798-A.
XX
XX PD 22-MAR-1990.
XX
XX PF 02-SEP-1988; 88US-00240856.
XX
XX XX 02-SEP-1988; 88US-00240856.
XX
XX PR 24-JUL-1989; 89US-00384508.
XX
XX XX (GETH ) GENENTECH INC.
XX
XX FI Anderson S, Bennett WZ, Botstein D, Higgins DL, Paoni NF;
XX Zoller M;
XX
XX DR WPI; 1990-115987/15.
XX
XX PT Modified tissue plasminogen activator - activated only when proximate to
XX plasmin at site of clot and not systemically.
XX
XX PS Claim 17; Page 44; 63pp; English.
XX
XX CC The t-PA variants represented in AAR09215-63 (and the T252R or N184S
XX analogues thereof or combinations thereof) and AAR09266-83 have their
XX fibrinolytic activity turned off when generally in the plasma and
XX activated when proximate to plasmin at the site of the clot thus
XX providing specific localised clot therapy
XX
XX SQ Sequence 439 AA;
XX
XX Query Match 25.2%; Score 794; DB 2; Length 439;
XX Best Local Similarity 37.4%; Pred. No. 1e-39;
XX Matches 176; Conservative 68; Mismatches 166; Indels 60; Gaps 15;
XX
QY 112 QNTCKDNPCGRCGLITQSPPIYR-----CVCKHPYTPSCSQVVPV--CRPNPC 159
DB 3 QVICRDE---KTQMIYQHQSLRPLVLRNRYEYWCN---SGRAQCHSVPVKSCSEPRC 56
QY 160 QNGATCSRHKRSRFTACAPQPKGKCEIGS-DDCVVGDGYSYRGKMTVQHQACLYW 218
DB 57 FNGTCCQALYFSDVFCQCPGFAKCCEDTRATCFNGSAVGRHSTESGASCLPW 116
QY 219 NSHLLQENYMFMDAETHGIGEHNFNRPDADEKPCWFIKVTNDKVKWEYCDVSACSA 278
DB 117 NSMILRARRYTAQNPSAQAALGLGKHNYCRNPDGDAKPCHV-LKRRLLTWEYCDVPSCS- 174
QY 279 QDVAYPEESPTEPTKLPDPSCKTETAEKIKRIYGGFKSTAGKHPWCASLQSSPLT 338
DB 175 -----TCGLRQYSQPF-RIKGGLFADIASHPWQAALPAK--HR 210
QY 339 ISMPQGHFCGALHPWVLPAACTD--IKTRHLKVLGDDQDLKKEEFHEQSFVREKIF 396
DB 211 RSPGERHLCGILISSCWILSAACHFQERPPPHLTVILGRTYRVVPGEEQKFEVKYI 270
QY 397 KYSHNERDEIHPNDIALKLPVDGHCALESKYVKTVCIP--DGSPSPSSECHI SOWGV 454
DB 271 VHKFED--DDTYNDIALQLKSSRCAQESSVVRTVCLPPADLPDWTCELSGYGK 328
QY 455 TETGKG--SROLLDAKVLIAINTLCNSRQLYDHMDISMCAGNLQKFG-----QDTCQG 507
DB 329 HEALSPFYSERLKEAHVRLYFSSRCTSQHLNLRVTVDNMLCAGDTRSGGQANLHDAQCG 388
QY 508 DSGGLPCEKDGTYVVGIVSWGLECGKR--PGYVTVQTKFLNWKATIK 555
DB 389 DSGGFLVCLNDGRMTLVGIIISWGLGCGQKQVPGYVTKVTVLYLDWRDNR 438

```

RESULT 14
AAP71742

```

ID AAP71742 standard; protein; 623 AA.
XX
XX AC AAP71742;
XX
XX DT 07-JUN-1991 (first entry)
XX
XX DE Ile(277)t-PA with extra finger- and extra growth factor-domains.
XX
XX KW tissue plasminogen activator; finger domain; fibrinolysis.
XX
XX OS Synthetic.
XX
XX FH Key Location/Qualifiers
XX Domain 9..46
XX FT /label= extra finger domain
XX FT 54..87
XX FT /label= extra growth factor domain
XX FT Protein 97..623
XX FT /label= natural human tPA
XX FT /note= "Lys(277) replaced by Ile"
XX
XX PN EP241210-A.
XX
XX XX 14-OCT-1987.
XX
XX PF 01-APR-1987; 87EP-00302837.
XX
XX PR 02-APR-1986; 86GB-00008020.
XX
XX PA (BEEC ) BEECHAM GROUP PLC.
XX
XX PI Browne MJ, Robinson JH;
XX
XX DR WPI; 1987-286054/41.
XX
XX PT Modified fibrinolytically active plasminogen activator - has added finger
XX domain(s) for antithrombotic use and is obcd. by recombinant DNA
XX procedures.
XX
XX PS Claim 11; Page 12; 16pp; English.
XX
XX CC The modified tPA molecules retain the fibrinolytic activity of natural
XX tPA and are used to treat thrombotic diseases. The extra finger domain
XX extends at the N-terminus to comprise amino acids 1-5 of native tPA. The
XX tPA mutein can form part of a hybrid protein when linked to another
XX fibrinolytically active protease so that the catalytic site of the hybrid
XX protein essential for fibrinolytic activity is optionally blocked by a
XX removable blocking group. See also AAP71740-1 and AAP71743-5
XX
XX SQ Sequence 623 AA;
XX
XX Query Match 25.1%; Score 793; DB 1; Length 623;
XX Best Local Similarity 31.2%; Pred. No. 1.6e-39;
XX Matches 188; Conservative 75; Mismatches 182; Indels 158; Gaps 17;
XX
QY 77 CQNPCEHGGDC--LVHGSTFTCSCLAPFSGNKQ-----KVQNTCKDNPCGRCGL 126
DB 54 CSEPCFNGGTCCQALYFSDVFCQCPGFAKCCEDTRATPFGSVQVICRDE---KTQMI 110
QY 127 ITQSPPIYR-----CVCKHPYTPSCSQVVPV--CRPNPCQNGATCSRHKRSKF 174
DB 111 YQHQSLRPLVLRNRYEYWCN---SGRAQCHSVPVKSCSEPRCFNGTCCQALYFSDF 167
QY 175 TCACPDQPKGKFCET-----
DB 168 VQCPGEGAGKCCEDTRATCYEDGIGSYRGTWSTAESGABCTNWNSSALQAPKYSGRRP 227
QY 190 -----GSDCYVGDGYSYRGK 205
DB 228 DAIRLGLGNHNYCRNPDSDKPCWCVFKAGYSSEFCSTPACSEGNSDCYFGNSAYRGT 287
QY 206 MNRTVQHQACLYWNSHLLQENYMFMDAETHGIGEHNFNRPDADEKPCWFIKVTNDK 265

```

Db 288 HSLTESGASCLPWNMILIGKYYTAQNPQAALGLGKHNVCENPDGDAKFWCHV-LKNRR 346
Qy 266 VKWEYCDVSACSADQVAYPEESPTFSTKLPDFDSCGKTEIAERKIKIYGGFKSTAGKH 325
Db 347 LTWEYCDVPSCS-----TCGLRQVSPQF-RIIGLFAFIASH 383
Qy 326 PWQASLOSPLTISMPQGHFGGALIHPCWVLTAAHCTD--IKTRHLKVLGDDQLKKE 383
Db 384 PWQAIFAK--HRRSPGERFLCGGILISSCWILSAHKFQERPPPHLTVILGRTYRVP 441
Qy 384 EFHEQSFVEKIPKYSHYNERDEIPHNDIALKLKPVGDHCALESKYVTKVCLP--DGSF 441
Db 442 GEEQKFEVEKIVHKEFD--DDTYDNDIALQLKSDSSRCAQESSVVRTVCLPPADLQ 499
Qy 442 PGSECHISGWVTETGKG--SRQLLDKAVKLANTLNSRQLYDHMIDDSMICAGNLQK 499
Db 500 PDWTECELSGKGHEALSFFYERLKEAHRVLPSSRCTSQHLLNRTVTDNMLCAGDTRS 559
Qy 500 PG-----QDTCCGDSGGPLTCKDGTYYVYGVISWGLECGKR--PGVYTVTKFLNWIKA 552
Db 560 GGPOANLHDACQDGGGGLVCLNDGRMTLVGIISWGLGCGQKDVPGVYTKVTNYLDWIRD 619
Qy 553 TIK 555
Db 620 NMR 622

RESULT 15
ID AAP71740
AC AAP71740; standard; protein; 623 AA.
XX
XX
XX 07-JUN-1991 (first entry)
XX
XX t-PA with extra finger domain and extra growth factor domain.
XX tissue plasminogen activator; finger domain; fibrinolysis.
XX Synthetic.
XX
XX Key Location/Qualifiers
FH Domain 9..46
FT Domain /label= extra finger domain
FT Domain 54..87
FT /label= extra growth factor domain
FT Protein 97..623
FT /label= natural human tPA
XX EP241210-A.
XX
XX
XX 14-OCT-1987.
XX
XX 01-APR-1987; 87EP-00302837.
XX
XX 02-APR-1986; 86GB-00008020.
XX
XX (BEEC) BEECHAM GROUP PLC.
XX
XX Browne MJ, Robinson JH;
XX
XX WPI; 1987-286054/41.
XX
XX Modified fibrinolytically active plasminogen activator - has added finger
XX domain(s) for antithrombotic use and is obtd. by recombinant DNA
XX procedures.
XX
XX
XX Claim 11; Page 12; 16pp; English.

CC The modified tPA molecules retain the fibrinolytic activity of natural
CC tPA and are used to treat thrombotic diseases. The extra finger domain
CC extends at the N-terminus to comprise amino acids 1-5 of native tPA. The
CC tPA mutein can form part of a hybrid protein when linked to another

CC fibrinolytically active protease so that the catalytic site of the hybrid
CC protein essential for fibrinolytic activity is optionally blocked by a
CC removable blocking group. See also AAP71741-5
XX
SQ Sequence 623 AA;
Query Match 25.1%; Score 792; DB 1; Length 623;
Best Local Similarity 31.2%; Pred. No. 1.9e-39;
Matches 188; Conservative 75; Mismatches 182; Indels 158; Gaps 17;
Qy 77 CQNPCEHGGDC--LVHGSTFTCSCLAPFSNGKCO-----KVONTCKNDPCCGRGQCL 126
Db 54 CSBPCFNGGTCQOALYFSDFCVCCPEGFAGKCCIDTRATPGSYQVCRDE---RTOMI 110
Qy 127 ITQSPPYR-----CVCKHPYTPGSCSQWVPV--CRPNPCONGATCSRHKRSKF 174
Db 111 YQHQSWLRVLRNSRVEYWCN---SGRAQCHSVFVKSCSEPRCFNGGTCQOALYFSD 167
Qy 175 TCACPDQFKKFCBI-----GSDDCYVGDGYSYRGK 205
Db 168 VCQCPGCFAGKCCIDTRATCYEDQGISYRGWTSTAESGAECTNMNSSALLAQKPYSGRRP 227
Qy 190 -----GSDDCYVGDGYSYRGK 205
Db 228 DAIRLGLGNHNYCRNDRDSKPCWYVFKAGKYSSEFCSTPACSEGNSDCYFGNGSAYRG 287
Qy 206 MNRVTNQHACLYWNHLLQENYNNFMEDAETHIGIEHNFCRNPDADEKPCFIKVTNDK 265
Db 288 HSLTESGASCLPWNMILIGKYYTAQNPQAALGLGKHNVCENPDGDAKFWCHV-LKNRR 346
Qy 266 VKWEYCDVSACSADQVAYPEESPTFSTKLPDFDSCGKTEIAERKIKIYGGFKSTAGKH 325
Db 347 LTWEYCDVPSCS-----TCGLRQVSPQF-RINGGLFADIASH 383
Qy 326 PWQASLOSPLTISMPQGHFGGALIHPCWVLTAAHCTD--IKTRHLKVLGDDQLKKE 383
Db 384 PWQAIFAK--HRRSPGERFLCGGILISSCWILSAHKFQERPPPHLTVILGRTYRVP 441
Qy 384 EFHEQSFVEKIPKYSHYNERDEIPHNDIALKLKPVGDHCALESKYVTKVCLP--DGSF 441
Db 442 GEEQKFEVEKIVHKEFD--DDTYDNDIALQLKSDSSRCAQESSVVRTVCLPPADLQ 499
Qy 442 PGSECHISGWVTETGKG--SRQLLDKAVKLANTLNSRQLYDHMIDDSMICAGNLQK 499
Db 500 PDWTECELSGKGHEALSFFYERLKEAHRVLPSSRCTSQHLLNRTVTDNMLCAGDTRS 559
Qy 500 PG-----QDTCCGDSGGPLTCKDGTYYVYGVISWGLECGKR--PGVYTVTKFLNWIKA 552
Db 560 GGPOANLHDACQDGGGGLVCLNDGRMTLVGIISWGLGCGQKDVPGVYTKVTNYLDWIRD 619
Qy 553 TIK 555
Db 620 NMR 622

Search completed: May 24, 2004, 09:46:50
Job time : 56 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 24, 2004, 09:43:55 ; Search time 16.5 Seconds
(without alignments)
3264.682 Million cell updates/sec

Title: US-09-912-559-3

Perfect score: 3154

Sequence: 1 MFARMSDLHLLMALVGKT.....TQVTKFLNWKATIKSBSGF 560

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 segs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 78: *
1: PIR1: *
2: PIR2: *
3: PIR3: *
4: PIR4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3154	100.0	560	1 JC4795	plasma hyaluronan-
2	2528	80.2	558	2 JC5978	plasma hyaluronan-
3	832.5	26.4	555	1 A46888	hepatocyte growth
4	773.5	24.5	603	2 S28941	coagulation factor
5	766.5	24.3	562	1 UKHUT	t-plasminogen acti
6	757.5	24.0	559	1 A35029	t-plasminogen acti
7	745.5	23.6	559	1 A29941	t-plasminogen acti
8	741	23.5	615	1 KFHU12	coagulation factor
9	731	23.2	477	1 A34369	t-plasminogen acti
10	731	23.2	477	2 JS0598	t-plasminogen acti
11	720.5	22.8	431	2 JS0599	t-plasminogen acti
12	706.5	22.4	593	2 S45281	coagulation factor
13	704.5	22.3	477	2 JS0597	t-plasminogen acti
14	704	22.3	432	1 S18932	u-plasminogen acti
15	686	21.8	442	1 UKPG	u-plasminogen acti
16	683.5	21.7	433	1 JN0560	u-plasminogen acti
17	680	21.6	433	1 UKMS	u-plasminogen acti
18	660	20.9	433	1 UKBAY	u-plasminogen acti
19	660	20.9	812	1 PLBO	plasma (EC 3.4.21
20	658.5	20.9	431	1 UKHU	u-plasminogen acti
21	650.5	20.6	434	1 A35005	u-plasminogen acti
22	643.5	20.4	394	2 JS0600	u-plasminogen acti
23	642.5	20.4	810	1 PLHU	plasma (EC 3.4.21
24	635.5	20.1	810	2 B30848	plasma (EC 3.4.21
25	607	19.2	790	1 PLPG	plasma (EC 3.4.21
26	598.5	19.0	812	1 PLMS	apoprotein(a) (EC
27	592.5	18.8	4548	1 SC0657	plasma (EC 3.4.21
28	587	18.6	455	2 A61545	plasma (EC 3.4.21
29	578.5	18.3	1420	2 A32869	apolipoprotein(a)

30	571.5	18.1	810	2 I46260	plasma (EC 3.4.21
31	557	17.7	460	2 B61545	plasma (EC 3.4.21
32	467.5	14.8	475	1 EXCH	coagulation factor
33	465	14.7	716	1 JC5061	macrophage-stimula
34	462.5	14.7	761	2 JC5759	brain-specific ser
35	461	14.6	716	1 A40332	macrophage-stimula
36	457.5	14.5	728	1 A60185	hepatocyte growth
37	456.5	14.5	728	1 A35644	hepatocyte growth
38	444	14.1	710	1 IS1283	coagulation factor
39	439.5	13.9	461	1 KFHU	macrophage-stimula
40	438.5	13.9	711	1 A47136	membrane-bound arg
41	438.5	13.9	855	2 JC7731	thrombin (EC 3.4.2
42	437.5	13.9	625	1 TBBO	thrombin (EC 3.4.2
43	429	13.6	622	1 TBHU	coagulation factor
44	427	13.5	482	1 EXRT	hepatocyte growth
45	423.5	13.4	728	1 JH0579	hepatocyte growth

ALIGNMENTS

RESULT 1

JC4795
plasma hyaluronan-binding protein precursor - human
N:Alternate names: hepatocyte growth factor activator-like protein; PHBP
N:Contains: serine proteinase (EC 3.4.21.-)
C:Species: Homo sapiens (man)
C:Date: 15-Oct-1995 #sequence_revision 16-Aug-1996 #text_change 19-Jul-2002
C:Accession: JC4795
J:Choi-Miura, N.H.; Tobe, T.; Sumiya, J.; Nakano, Y.; Sano, Y.; Mazda, T.; Tomita, M.
J: Biochem. 119, 1157-1165, 1996
A:Title: Purification and characterization of a novel hyaluronan-binding protein (PHBP)
X activator.
A:Reference number: JC4795; MUID:96425001; PMID:8827452
A:Accession: JC4795
A:Molecule type: mRNA
A:Residues: 1-560 <CHO>
A:Cross-references: NID:G1836158; PIDN:AA846909.1; PID:G1836159
A:Experimental source: plasma
A:Note: parts of this sequence, including the amino ends of the mature chains, were deter
C:Genetics:
A:Gene: GDB:HABP2; HABP; PHBP; HGFAL
A:Cross-references: GDB:4573962
C:Complex: a disulfide-bonded heterodimer of chains produced from the same precursor; th
C:Superfamily: plasma hyaluronan-binding protein; EGF homology; kringle homology; trypsi
C:Keywords: chondroitin sulfate proteoglycan; glycoprotein; hyaluronic acid; hydrolase;
F:1-23/Domain: signal sequence #status predicted <SIG>
F:24-313/Product: plasma hyaluronan-binding protein, 50K chain #status predicted <50K>
F:77-108/Domain: EGF homology <EG1>
F:115-147/Domain: EGF homology <EG2>
F:194-276/Domain: kringle homology <KRI>
F:314-516/Product: trypsin homology <TRY>
F:54,207/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:77-88,82-97,99-108,115-125,120-136,138-147,154-165,159-176,178-187,194-276,215-257,246
F:362,405,509/Active site: His, Asp, Ser #status predicted

Query Match 100.0%; Score 3154; DB 1; Length 560;

Best Local Similarity 100.0%; Pred. No. 3e-188;

Matches 560; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MFARMSDLHLLMALVGKTACGFSLSLLESIDPDWTPDQYDYSYEDYNOENTSSLT 60

DB 1 MFARMSDLHLLMALVGKTACGFSLSLLESIDPDWTPDQYDYSYEDYNOENTSSLT 60

QY 61 HAENPDWYTEDQADPCQPNPCHEGGDCLVHGSTFTTCSCLAPFGSKCKQVQNTCKDNPC 120

DB 61 HAENPDWYTEDQADPCQPNPCHEGGDCLVHGSTFTTCSCLAPFGSKCKQVQNTCKDNPC 120

QY 121 GRGCLITQSPPYRYRCVKHPYTGPCSQVVPVCRPNPCQNGATCSRHKRSKFTACPD 180

DB 121 GRGCLITQSPPYRYRCVKHPYTGPCSQVVPVCRPNPCQNGATCSRHKRSKFTACPD 180

181	QY	QFKGFCIEIGDDCYVGDGYSYRGKMRTYNQHACLYWNSHLLLOENYNMFWEDEATHGI	240
181	Db	QFKGFCIEIGDDCYVGDGYSYRGKMRTYNQHACLYWNSHLLLOENYNMFWEDEATHGI	240
241	QY	GEHNFRCRNPDADEKPCWCFIKVTNNDKVXWEYCDVSACSADQVAYPEESPEPSTKLPGGDS	300
241	Db	GEHNFRCRNPDADEKPCWCFIKVTNNDKVXWEYCDVSACSADQVAYPEESPEPSTKLPGGDS	300
301	QY	CGKTEIABERKIKRYGGFKSTAGKHPPQASLQSSPLTITSMPOGHFCGGALHPCWILTA	360
301	Db	CGKTEIABERKIKRYGGFKSTAGKHPPQASLQSSPLTITSMPOGHFCGGALHPCWILTA	360
361	QY	AHCTDIKTRHLKVLVJGDODLKKBEFFHQSPRVSKIFKYSHYNDRDIPENDIALLKLPV	420
361	Db	AHCTDIKTRHLKVLVJGDODLKKBEFFHQSPRVSKIFKYSHYNERDEIPENDIALLKLPV	420
421	QY	DGHCALESKYVKTVCLPDGSPFSGSECHIISGWTETGKSGRQLLDAKVKLIANTLCNSR	480
421	Db	DGHCALESKYVKTVCLPDGSPFSGSECHIISGWTETGKSGRQLLDAKVKLIANTLCNSR	480
481	QY	QLYDHWIMDSMICAGNLQPKQDTCQDSDSGPLTCEKDGTYYYVYGIVSGLECGKEPGYV	540
481	Db	QLYDHWIMDSMICAGNLQPKQDTCQDSDSGPLTCEKDGTYYYVYGIVSGLECGKEPGYV	540
541	QY	TQVTKPLANWIKATIKSESGF	560
541	Db	TQVTKPLANWIKATIKSESGF	560

RESULT 2

JC5878
plasma hyaluronan-binding protein precursor - mouse
C;Species: Mus musculus (house mouse)
C;Date: 11-Mar-1998 #sequence_revision 11-Mar-1998 #text_change 16-Jul-1999
C;Accession: JC5878
R;Hashimoto, K.; Tobe, T.; Sumiya, J.; Saguchi, K.; Sano, Y.; Nakano, Y.; Choi-Miura, N.
Biol. Pharm. Bull. 20, 1127-1130, 1997
A;Title: Cloning of the cDNA for a mouse homologue of human PHBP: A novel hyaluronan-binding protein precursor from mouse.
A;Reference number: JC5878; MUID: 98052539; PMID: 9401717
A;Accession: JC5878
A;Molecule type: mRNA
A;Residues: 1-558 <HAS>
C;Comment: This protein acts as serine protease.
C;Superfamily: plasma hyaluronan-binding protein; EGF homology; kringle homology; trypsin
F;1-23/Domain: Signal sequence #status predicted <SIG>
F;24-311/Product: plasma hyaluronan-binding protein large chain #status predicted <MATL>
F;75-106/Domain: EGF homology <EG1>
F;113-145/Domain: EGF homology <EG2>
F;152-185/Domain: EGF homology <EG3>
F;192-274/Domain: kringle homology <KR1>
F;312-558/Product: plasma hyaluronan-binding protein small chain #status predicted <MATS>
F;312-548/Domain: trypsin homology <TRY>

QY 241 GHNFCNPNDADEKWCFFTKVNDKVKWBYCDVSACSADVAYPEESPTPEPTKLPGLPDS 300
 DB 239 AEHNFCNPDGDKPWCFFVKYSEKVKWBYCDVTVCPVDPDTPNPVESLLEPVKMLPGPES 298
 QY 301 CGKTEIAERKIKIYGGFSTAGKHPWQASLOSSIPLTISMFGQHC CGGALHPCHWLTA 360
 DB 299 CGKTEVAEHAIVKIKIYGGFSTAGKHPWQVSLQTSPLTSMFGQHC CGGALHPCHWLTA 358
 QY 361 AHCTDIKTRHLKVVVLGDQDLKKEEFHEQSFVREKIFKYSHYNERDEIPHNIDIALLKLPV 420
 DB 359 AHCTDINTKHLKVVVLGDQDLKKTSTESHEQTPRVEKILKYQYNERDEIPHNIDIALLKLPV 418
 QY 421 DGHCALESKVYKIVCIPLDGSFSGSECHISGKGVTTGSGRQLLDAKVLIANTICNSR 480
 DB 419 GGHCALESRVYKIVCIPLDPSDFPFGSGTCHISGKGVTEITGSGRQLLDAKVLIANPLCNSR 478
 QY 481 QLYDHMIDDSMICAGNLQKPGQDTCQDGGPPTCCKDGTYYVYGVYVSMGLECGKPKPGVY 540
 DB 479 QLYDHTIDDSMICAGNLQKPGSDTCQDGGPPTCCKDGTYYVYGVYVSMGLECGKPKPGVY 538
 QY 541 TQVTKFLNWIATIKSESQ 559
 DB 539 TQVTKFLNWIKTWHREAG 557
 RESULT 3
 A46688
 hepatocyte growth factor activator (EC 3.4.21.-) precursor [validated] - human
 C:Species: Homo sapiens (man)
 C:Date: 21-Sep-1993 #sequence_revision 25-Aug-1995 #text_change 08-Dec-2000
 C:Accession: A46688
 R:Mayazawa, K.; Shimomura, T.; Kitamura, A.; Kondo, J.; Morimoto, Y.; Kitamura, N.
 J. Biol. Chem. 269, 10024-10028, 1993
 A:Title: Molecular cloning and sequence analysis of the cDNA for a human serine proteas-
 e d coagulation factor XII.
 A:Reference number: A46688; MUID:93252878; PMID:7683665
 A:Accession: A46688
 A:Molecule type: mRNA
 A:Residues: 1-655 <MID>
 A:Cross-references: DDBJ:D14012; NID:g219680; PIDN:BAA03113.1; PID:g219681
 A:Experimental source: liver (mRNA); serum (protein)
 A>Note: sequence extracted from NCBI backbone (NCBIN:131227, NCBIPI:131228)
 A>Note: parts of the sequence, including the amino ends of the heavy and light chains,
 C:Genetics:
 A:Gene: GDB:HGPAC; HGFA; HGFAP
 A:Cross-references: GDB:9954514
 A:Map position: 4p16-4p16

C:Function: activates hepatocyte growth factor by specific proteolytic cleavage
A:Pathway: tissue repair and regeneration
C:Superfamily: coagulation factor XII; EGF homology; fibronectin type I repeat homology
C:Keywords: glycoprotein; hydrolase; kringle; liver; plasma; serine proteinase
F:1-34/Domain: signal sequence #status predicted <SIG>
F:108-148/Domain: fibronectin type II repeat homology <LF2>
F:164-197/Domain: EGF homology <EG1>
F:202-237/Domain: fibronectin type I repeat homology <LF1>
F:245-278/Domain: EGF homology <EG2>
F:286-367/Domain: kringle homology <KR>
F:373-407/Product: hepatocyte growth factor activator light chain #status experimental
F:408-655/Product: hepatocyte growth factor activator heavy chain #status experimental
F:408-641/Domain: trypsin homology <TR>
F:40-48,290,468,492,546/binding site: carboxydrate (Asn) (covalent) #status predicted
F:164-175,169-186,188-197,202-230,228-237,245-256,250-267,269-278,286-367,307-349,338-347,407,508/active site: His, Asp, Ser #status predicted

QY 75 -----DPCPNPCEHGGDC--LVHGSTFTSCSLAPFSGNKKQKQVQNTCKDNKPCGRCQCL 126
DB 156 GPPAALDFCAGPCLNGSCSNTQDPOSHYSCSPRAFTG-----KD--CGTEKCF 203
QY 127 ITQSPYY-----RCVCKHPVTGSCSQVV--PVCRRNPONGATCSRHK 169
DB 204 DETRYELEGDRWARVRQHVQCEC---FGGTWCEGTRHACLSPLCLNGTCHLIV 260
QY 170 RRSFTACAPDQFGKFCFCEIGSD- CYVGDGYSYRGKRNRTVNOHACLYWNSHLLQENY 228
DB 261 ATGTTVCACPPGAGRLCNIEPDERCFLNGTGYRGVASTSASGLSCLAWNSDILLYQELH 320
QY 229 NMFMEDAETHGIEHNFGRNPDADKPCFKIVTNDKVKEYCDVSACS--QDVAYPEE 286
DB 321 VDSVCAALLGLGHAYCRNPDNDRPWCYV-VKDSALSWEYCLBACESLTRVQLSPDL 379
QY 287 SPTPESTKLPDFDSCGCKTEIAERKIK-RIYGGFKSTAGKHFWQASLOSSLPITISMPCQH 345
DB 380 LATLPEPASPRQACGRHRKRTFLRPRIIGSSSLPGSHPWLAIVIG-----DS 430
QY 346 FCGGALIHPCWVLTAAHCTDIKTRH--LKVVLGQDLKKEFBHQSRVEKIFKYSYHNE 403
DB 431 FCAGSLVHTCWVYSAHCFSHSPRDSVSVVLGQHFNRRTDVTQTFGIEKIYIPTYLSV 490
QY 404 RDEIPHNDIALKLPVDGHCALESKYVKTVCCLPD--GSFPSSGECHISGWG-VTETGKG 460
DB 491 FNPDSH-DLVLRLLKKGDRCATRSQFVQPCLEPEPGSTFPAGHKCQIAGWGLDENVSG 549
QY 461 -SRQLLDARVXLANTLNSRLQDMDHIDMSICAGNLQKPGQDTCQDSDSGPPTCEKDG 519
DB 550 YSSSLREALVPLVADHKCSPEVYAGDISPNMLCAGYDFCK-SDACQDSDSGPLACEKNG 608
QY 520 TTYVYGVISWGLECKK--RPGVYTVQTKFLNWKATIK 555
DB 609 VAYLYGISWGDGCGRLHKPGVITRVANYVDWINDRIR 646

RESULT 4

S28941
coagulation factor XIIa (EC 3.4.21.38) - guinea pig (fragment)
N:Alternate names: Hageman factor
C:Species: Cavia porcellus (guinea pig)
C:Date: 25-Feb-1994 #sequence_revision 03-Aug-1995 #text_change 21-Jan-2000
C:Accession: S28941
R:Semba, U.; Yamamoto, T.; Kunisada, T.; Shibuya, Y.; Tanase, S.; Kanbara, T.; Okabe, H.
Biochim. Biophys. Acta 1159, 113-121, 1992
A:Title: Primary structure of guinea-pig Hageman factor: sequence around the cleavage site
A:Reference number: S28941; MUID:93003367; PMID:1390917
A:Accession: S28941
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-603 <SEM>
A:Cross-references: EMBL:X68615; NID:G49578; PIDN:CAA48600.1; PID:G49579
C:Superfamily: coagulation factor XII; EGF homology; fibronectin type I repeat homology
C:Keywords: hydrolase; serine proteinase
F:46-87/Domain: fibronectin type II repeat homology <F1>
F:134-169/Domain: fibronectin type I repeat homology <F2>
F:177-208/Domain: EGF homology <EGF>
F:216-294/Domain: kringle homology <KR>
F:359-597/Domain: trypsin homology <TRY>
Query Match 24.5%; Score 773.5; DB 2; Length 603;
Best Local Similarity 34.1%; Pred. No. 1e-40;
Matches 208; Conservative 74; Mismatches 205; Indels 123; Gaps 29;
QY 17 VGTACGFLMSLLESIDPDWTPDQDYSDYEDYNQBEEN-----TSSTLTHAENPDWY 69
DB 37 VGLTVTG-----EPCYFFPQYRQLYHCHIKGRGPRPWCATTNFPDQDQWAY 86
QY 70 ---TEDQADPC-QPNPCEHGGDCLVHGSTFTSCSLAPFSGNKKQKQVQNTCKDNKPCGRCQCL 125
DB 87 CLBPKVKVDKCSKHPNCPQRCGICVNTLSLSPHCLCPDHLTKGHCQ-----REKC 134

QY 126 LITQSPYY-----RCVCKHPVTGSCSQV--PVCRRNPONGATCSRHK 169
DB 135 FEPQJHRRFFHENEIWRTPGPAVAKCHKGP--DAHCKQMHSSQSQGTNPCLNGGRCL--E 190
QY 170 RRSKFTACAPDQFGKFCFCEIGSD-DCYVGDGYSYRGKRNRTVNOHACLYWNSHLLQENY 228
DB 191 VEGHILCDPCMGYGTGPFCDLDTTASCYBGRGYSYRGMARITTVSGAKCORWASEATYR--- 247
QY 229 NMFMEDAETHGIEHNFGRNPDADKPCFKIVTNDKVKEYCDVSACS--AQDVAYPEE- 285
DB 248 NMTAEQALRRGLGHTFCRNPNDTRPWCFTVMGN-RLSWEYCDLAQCYPPQPTATPHD 306
QY 286 --ESPTPESTK-----LPGFDS--CGKTEIAER--KIKRIYGGPKS 320
DB 307 REHPKLPSSRLSILQTPQTTQNALANELPETSSLLCGQ-RLRKRLLSSLSRIYVGLVA 365
QY 321 TAGKHFWQASLOSSLPITISMPCQHFCGALHPCWLTAAHCTDIK--TRHLKVLGQD 378
DB 366 LPGAHPYIAALYWG-----SNFCSGLIAPCWLTAAHCLQNRPAPELKVVLGQD 416
QY 379 DLKKEBFHQ-----SFRVEKIFKISHYNERDEIPHNDIALKL-KPVDGHCALESKYVK 432
DB 417 RHNQSCHEQCTLAVHSYRLHEAFSPSSY-----LNDLALLRLQKSDAGSCAQLSPYVQ 469
QY 433 TVCLPDG-SFPGSE---CHISGWGTETG--KGRQLLDKAVKLIANTLNSRLQYDEM 486
DB 470 TVCLPSGPAPPSESETTCCEVAGWGHQFEGAEYSFLOEAQVPLISSRCSPEVHGDA 529
QY 487 IDDSMLCAGNLQKPGQDTCQDSDSGPPTCEKDGTYV---VYGVISWGLECKGR--PGVYT 541
DB 530 FLSCMLCAFLF--GGTDACQDSDSGPLVCEDEAAEHLRLGIVSWGSCGDRNKEGVYT 588
QY 542 QVTKFLNWK 551
DB 589 DVASYLTWIQ 598

RESULT 5

UKHUT
t-plasminogen activator (EC 3.4.21.68) precursor [validated] - human
N:Alternate names: t-PA; tissue plasminogen activator
C:Species: Homo sapiens (man)
C:Date: 14-Nov-1983 #sequence_revision 14-Nov-1983 #text_change 08-Dec-2000
C:Accession: A94004; A23529; J70562; A93293; S02125; A91343; A93951; A91322; A54645; I6C
R:NY, T.; Eigh, F.; Lund, B.
Proc. Natl. Acad. Sci. U.S.A. 81, 5355-5359, 1984
A:Title: The structure of the human tissue-type plasminogen activator gene: correlation
A:Reference number: A94004; MUID:84298137; PMID:6089198
A:Accession: A94004
A:Molecule type: DNA
A:Residues: 1-562 <NVT>
A:Cross-references: GB:I00141
A:Note: the codon given for residue 93 (ACC) is inconsistent with the authors' translation
R:Frieznher Degen, S.J.; Rajput, B.; Reich, E.
J. Biol. Chem. 261, 6972-6985, 1986
A:Title: The human tissue plasminogen activator gene.
A:Reference number: A23529; MUID:86136143; PMID:3009482
A:Accession: A23529
A:Molecule type: DNA
A:Residues: 1-562 <DEG>
A:Cross-references: GB:X03021; NID:G339817; PIDN:AAA98809.1; PID:G339818
R:Itagaki, Y.; Yasuda, H.; Morinaga, T.; Mitsuda, S.; Higashio, K.
Agric. Biol. Chem. 55, 1225-1232, 1991
A:Title: Purification and characterization of tissue plasminogen activator secreted by
A:Reference number: J70562; MUID:91291340; PMID:1368681
A:Accession: J70562
A:Molecule type: mRNA
A:Residues: 31-562 <ITA>
A:Cross-references: DDBJ:D01096; NID:G220128; PIDN:BAA00881.1; PID:G441174
A:Experimental source: embryonic lung fibroblast IWR-90 cells
A:Note: part of this sequence, including the amino end of the mature protein, was confirmed
R:Pernica, D.; Holmes, W.E.; Kohr, W.J.; Harkins, R.N.; Vehar, G.A.; Ward, C.A.; Bennett

1007 OCT 19 1967

1
 2
 3
 4
 5
 6
 7
 8

2.

DB 402 TYNDIALQLKSSRCQAQESSVVRVCLPPADLQLPDWTECELSGYGKHEALSPYSE 461
QY 463 QLLDAKVLIANTLCNRQLYDHMDSDMICAGNLKQPG-----QDTCQDSGSGPLTCEK 517
DB 462 RLKEAHLVRLYPSRSTSHLLNRTVTDNMLCAGTRSGGQFQANJHDACQDSGSGPLVCLN 521
QY 518 DGTVVYGVISWGLECGKR--PGVTVQVTKFLNWKATIK 555
DB 522 DGRMTLVGIIISWGLCCGQKQVPGVYTKVNYLNDIRNMR 561
RESULT 6
A35029
t-plasminogen activator (EC 3.4.21.68) precursor - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: A35029; A31597
R:RefSeq, P.; Ohlsson, M.; Ny, T.
J. Biol. Chem. 265, 2022-2027, 1990
A:Title: The structure of the TATA-less rat tissue-type plasminogen activator gene. Spec
A:Reference number: A35029; MUID:90130448; PMID:2105315
A:Accession: A35029
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-559 <FEN>
A:Cross-references: GB:M31197; NID:G207429; PIDN:AAA42261.1; PID:G207431; GB:J05226
R:NY, T.; Leonardsson, G.; Hsueh, A.J.W.
DNA 7, 671-677, 1988
A:Title: Cloning and characterization of a cDNA for rat tissue-type plasminogen activator
A:Reference number: A31597; MUID:89170114; PMID:3148445
A:Accession: A31597
A:Molecule type: mRNA
A:Residues: 1-379, 'K', 381-559 <NYT>
A:Cross-references: GB:M23697; NID:G530159; PIDN:AAA41812.1; PID:G530160
C:Superfamily: tissue plasminogen activator; EGF homology; fibronectin type I repeat hom
C:Keywords: fibrinolysis; glycoprotein; hydrolase; kringle; serine proteinase
F:1-17/Domain: signal sequence #status predicted <SIG>
F:18-29/Domain: propeptide #status predicted <PRO>
F:30-559/Product: t-plasminogen activator #status predicted <MAT>
F:30-308/Product: t-plasminogen activator chain A #status predicted <ACH>
F:38-75/Domain: fibronectin type I repeat homology <1F1>
F:83-116/Domain: EGF homology <EGF>
F:124-205/Domain: kringle homology <KR1>
F:213-294/Domain: kringle homology <KR2>
F:309-559/Product: t-plasminogen activator chain B #status predicted <BCH>
F:309-553/Domain: trypsin homology <TRY>
F:338-68,66-75,83-94,88-105,107-116,124-205,145-187,176-200,213-294,234-276,265-289,297-4
F:149,481/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:308-309/Cleavage site: Arg-Ile (plasmin, trypsin) #status predicted
F:355,404,510/Active site: His, Asp, Ser #status predicted
Query Match 24.0%; Score 757.5; DB 1; Length 559;
Best Local Similarity 32.5%; Pred. No. 9.4e-40;
Matches 202; Conservative 86; Mismatches 192; Indels 141; Gaps 29;
QY 8 LHVLLMALVGKTACG--FSLMSLLSLDPDWTDPQDYD-----SYEDYNGEENTSS 57
DB 6 LCVLL-----CGVAFTL-----PDQIHRRFRGARSYATRCDEQTQT 45
QY 58 TLTHAENPDW---YYTEQADPCQNP-----CEHGDC--LVHGSTF 95
DB 46 --TYQQHOSWLRPMLRGNRVEYCRNCSGLAOCHSVVRSCSEPRCFNGGTCCQALYPSDF 103
QY 96 TCSCLAPSGNKKCQ-KVQNTCKDNPCGRGQCLITQSPYRVCVKHPYTGSPC----- 147
DB 104 VQCCPDGFGVKECDIDTRATCFE---GGQ---IT-----YRGWSTAENGAEINWNSSA 152
QY 148 --SQVFCVRPNPCONGATCSHKRSKFTACPDQ-----FK-GKF-----C 187
DB 153 LSKQPSYARRPNAIKLG--LGNHN-----YCRNPRDVKPMCYVFKAGKYTFEFCSTPAC 205
QY 188 EIG-SDDCVGDSYRGGMNRTVNOHACLYNWSHLLQLQENYVNMFMEDAEHGEHNC 246

DB 206 PKGPTEDCVGKGYRGTHSFTTSKASCLPWNMILIKTYTAWRANSQALGLGRHNYC 265
QY 247 RUPDADERKFWCIKNTDKVKWEYCDVSACSQDVAYPEESTPESTKLPGLGDSGKTEI 306
DB 266 RSPDGDGAPWCHV-NKDRKLTWEYCDMSPCS-----TCGLQKY 302
QY 307 AERKIKRIYGGFKSTAGKHPMQASLQSSILPLTISMPQGFQCGALHPCWLVTAACH--T 364
DB 303 KQOPF-RIKGGFTTITSHPMQAAI--FVGNKRSFGRFLCGGVLISSCWLSAAHCFVE 359
QY 365 DIKTRHLKVLGDQLKKEEFHEQSFVEKIFKYSHYNERDEIPNDTALLKLPVDGHC 424
DB 360 RFPFPHLKVVLGRTYRVVPGEEQTFEIKYTVHKEFD--DDTYDNDLALQLRSDSQ 417
QY 425 ALESKVKVTCVCLPDG--SPPSGSECHISQWGVYETGKG--SRQLLDKAVKLIANTLCNSR 480
DB 418 AQESSVGTACLPDQVQLPDWTECELSGYGKHEASSPFFSDRLKEAHVRLYPSRSTSQ 477
QY 481 QYDHMDSDMICAGNLKQPG-----QDTCQDSGSGPLTCEKDTYVYGVISWGLECGKR 536
DB 478 HLFNKTITSNMLCAGDTRTGGNQDVHDACQDSGSGPLVCMIDKRWTLGLIISWGLCGGOK 537
QY 537 --PGVTVQVTKFLNWKATIK 555
DB 538 DVPGIYTKVNYLNIQDNMK 558
RESULT 7
A22941
t-plasminogen activator (EC 3.4.21.68) precursor - mouse
C:Species: Mus musculus (house mouse)
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: A22941; S48205; S48207; S48206
J:Rickles, R.J.; Darrow, A.L.; Strickland, S.
J. Biol. Chem. 263, 1563-1569, 1988
A:Title: Molecular cloning of complementary DNA to mouse tissue plasminogen activator mR
A:Reference number: A22941; MUID:89087303; PMID:2826484
A:Accession: A22941
A:Molecule type: mRNA
A:Residues: 1-559 <R1C>
A:Cross-references: GB:J03520; NID:G202109; PIDN:AAA40470.1; PID:G202110
R:Lijnen, H.R.; van Hoef, B.; Beelen, V.; Collen, D.
Eur. J. Biochem. 224, 863-871, 1994
A:Title: Characterization of the murine plasma fibrinolytic system.
A:Reference number: S48202; MUID:95010076; PMID:7523120
A:Accession: S48205
A:Molecule type: protein
A:Residues: 33-37, 'X', 39-40 <LIJ>
A:Accession: S48207
A:Molecule type: protein
A:Residues: 309-316 <LI2>
A:Accession: S48206
A:Molecule type: protein
A:Residues: 33-37, 'X', 39-40 <LIW>
C:Superfamily: tissue plasminogen activator; EGF homology; fibronectin type I repeat hom
C:Keywords: fibrinolysis; glycoprotein; hydrolase; kringle; serine proteinase
F:1-17/Domain: signal sequence #status predicted <SIG>
F:18-29/Domain: propeptide #status predicted <PRO>
F:30-559/Product: t-plasminogen activator #status predicted <MAT>
F:30-308/Product: t-plasminogen activator chain A #status predicted <ACH>
F:38-75/Domain: fibronectin type I repeat homology <1F1>
F:83-116/Domain: EGF homology <EGF>
F:124-205/Domain: kringle homology <KR1>
F:213-294/Domain: kringle homology <KR2>
F:309-559/Product: t-plasminogen activator chain B #status predicted <BCH>
F:309-553/Domain: trypsin homology <TRY>
F:338-68,66-75,83-94,88-105,107-116,124-205,145-187,176-200,213-294,234-276,265-289,297-4
F:149,481/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:308-309/Cleavage site: Arg-Ile (plasmin, trypsin) #status predicted
F:355,404,510/Active site: His, Asp, Ser #status predicted
Query Match 23.6%; Score 745.5; DB 1; Length 559;

QY 196 VDDGYSYRGKMTNVTQVHACIYWNHLLQENYMFMDAETHGIGEHNFRCNPDADKP 255
DB 219 DGRGLSYRGLARTLTLSPGACQPFQWASEATYR---NVTAQARNWGLGGHAFRCNPDNIRP 275
QY 256 WCFIKVINDKVKWEYCDVSACSQAQVAYPESPT-----EPSTKL 295
DB 276 WCFV-LNRDRLSWEYCDLAQCTPQCAP---PTVPSPRLHVLMPAQAQPPKQPPIRT 331
QY 296 P-----GPDSCG-KTEIAERKIKRIYGGFKSTAGKHFWQASLQ 332
DB 332 PFSQOTPGALPAKREPPSLTRNGSLSCQGLRKSLSMTVRVGLVALRGAHYAALY 391
QY 333 SSLPLTISMPQGH-FCGGLIHPCWLTAAHCTDIK--TRHLKVLVGDQDLKKEEFHQSFRVEKIFKYS 384
DB 392 -----WGHSCFACSLIAPCWLTAAHCLQDRPAPEDLTIVLGOERRNHSCPCQT 441
QY 385 FHEQSFREYKIFKYSHYNERDEIPNDIALKL-KPVDGHCALESKYVTKVCLPDGSP-P 442
DB 442 LAVRSYRLHEAFSPVSY-----QHDLLALLRLQEDADGSCALLSPYVQVCLPSGAARP 494
QY 443 SGSE-CHISGCVGTGKSRQ-----LLDAKVKLIANTLCHNSRLQYDHMIDDSMICAGNL 497
DB 495 SETTLQVAGWG--HQFEGAEBYASFLQEAQVFPFSLERCSAPDVHGSILPGMLCAGFL 552
QY 498 QKPGQDTCQDGGSLTCEKDTG---YVYVIGVSWGLECGKR--PGVYTVQTKFLNMIK 551
DB 553 E-GGTDACQDGGSLVCEBQDAERRLT-LQGIISGSGCDRKNKPGVYTVDAVYLAWR 610

RESULT 9
A34369
t-plasminogen activator (EC 3.4.21.68) precursor - false vampire bat (Megaderma lyra)
C:Species: Megaderma lyra
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: A34369
R:Gardell, S.J.; Duong, L.T.; Diehl, R.E.; York, J.D.; Hare, T.R.; Register, R.B.; Jacob
J. Biol. Chem. 264, 17947-17952, 1989
A:Title: Isolation, characterization, and cDNA cloning of a vampire bat salivary plasmin
A:Reference number: A34369; MUID:90036867; PMID:2509450
A:Accession: A34369
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-477 <KRA>
A:Cross-references: GB:J05082; NID:gl66080; PIDN:AAA31596.1; PID:gl66081
C:Superfamily: tissue plasminogen activator; EGF homology; fibronectin type I repeat hom
C:Keywords: fibrinolysis; glycoprotein; hydrolyase; kringle; serine proteinase
F:1-21/Domain: signal sequence #status predicted <SIG>
F:22-36/Domain: propeptide #status predicted <PRO>
F:37-477/Product: plasminogen activator #status predicted <PLA>
F:42-79/Domain: fibronectin type I repeat homology <IFA>
F:87-120/Domain: EGF homology <EGF>
F:128-209/Domain: kringle homology <KRG>
F:226-471/Domain: trypsin homology <TRY>
F:42-72,70-79,87-98,92-109,111-120,128-209,149-191,180-204,214-345,257-273,265-334,359-4
F:185,398/Binding site: carbonylhydrolase (Asn) (covalent) #status predicted
F:225-226/cleavage site: His-Ser (plasmin) #status predicted
F:272,321,428/Active site: His, Asp, Ser #status predicted

Query Match 23.2%; Score 731; DB 1; Length 477;
Best Local Similarity 35.3%; Pred. No. 3.5e-38;
Matches 165; Conservative 73; Mismatches 171; Indels 58; Gaps 16;

QY 115 CKDNPGRGQCLITQSPPIYR-----CVCKHPYTPGSCSQVVPV--CRPNPCQNG 162
DB 42 CRDE---KTQMIYQQESMLRPEVRSKRVEHCRCDR---GLAQCHTVPVKSCSELRCFNG 95
QY 163 ATCSRHKRSKFTACPDQPKGKFCIGSD-DCYVGDGYSYRGKMTNVTQVHACLYWNH 221
DB 96 GTCWQAASFDFVCCPKGTGKQCEVDTHATCYKQGVYRGVTSSTSGAQCINWNSN 155
QY 222 LLLQENTYMFMDAETHGIGEHNFRCNPDADKPWCIFKVTNDKVKWEYCDVSACSQAQV 281
DB 156 LLTRRTYNGRRSDAITLGLGNHNYCRPNPNNSKPCVY-IKASKFILEFCSPVCSKA-- 212
QY 282 AYPESPTSTKLPFGDSCGKTEIAERKIKRIYGGFKSTAGKHFWQASLQSLPLTISM 341
DB 213 -----TCGLRKYPQLHST-GGLFTDITSHPWQAIFAQ--NRRSS 251
QY 342 PQGHFCGGLIHPCWLTAAHCTDIK--TRHLKVLVGDQDLKKEEFHQSFRVEKIFKYS 399
DB 156 LLTRRTYNGRRSDAITLGLGNHNYCRPNPNNSKPCVY-IKASKFILEFCSPVCSKA-- 212

QY 282 AYPESPTSTKLPFGDSCGKTEIAERKIKRIYGGFKSTAGKHFWQASLQSLPLTISM 341
DB 213 -----TCGLRKYPQLHST-GGLFTDITSHPWQAIFAQ--NRRSS 251
QY 342 PQGHFCGGLIHPCWLTAAHCTDIK--TRHLKVLVGDQDLKKEEFHQSFRVEKIFKYS 399
DB 252 GERFLCGGILLISSCWLTAAHCFQERYPPQHLRVVLGRVTVRKVPKGEQTFEVEKICIVHE 311
QY 400 HYNREDEIPNDIALKLKPEVDGHCALESKYVTKVCLPDGS--FPGSGSCHTSGMWGTET 457
DB 312 EFD--DDTYNNDIALQLKSGSPQCAQESDSVRAICLPEANLQLPDWTSCELSGYKHKHS 369
QY 458 GKG--SRQLLDAKVKLIANTLCHNSRLQYDHMIDDSMICAGNLQK-----PG-QDTCQDGG 510
DB 370 SSFPYSEQLKEGHVRLYPPSSRCTSKFLNKVTXKMLCAGDTRSGEIHNVHVDACQDGG 429
QY 511 GPLTCEKDTGYVYGVISWGLECGKR--PGVYTVQTKFLNMIKATIK 555
DB 430 GPLVCRNDHMTLLGIISGSGCGEKDIPGVYTKVTVNLGWRDNR 476

RESULT 10
J05098
t-plasminogen activator (EC 3.4.21.68) alpha-2 precursor - common vampire bat
N:Alternate names: tissue plasminogen activator
C:Species: Desmodus rotundus (common vampire bat)
C:Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 16-Jul-1999
C:Accession: J05098
R:Kaezschmar, J.; Haendler, B.; Langer, G.; Boidol, W.; Bringmann, P.; Alagon, A.; Dor
Gene 105, 229-237, 1991
A:Title: The plasminogen activator family from the salivary gland of the vampire bat Des
A:Reference number: J05097; MUID:92039036; PMID:1937019
A:Accession: J05098
A:Molecule type: mRNA
A:Residues: 1-477 <KRA>
A:Cross-references: GB:M63988; NID:gl66074; PIDN:AAA31593.1; PID:gl66075
C:Superfamily: tissue plasminogen activator; EGF homology; fibronectin type I repeat hom
C:Keywords: fibrinolysis; glycoprotein; hydrolyase; kringle; serine proteinase
F:1-21/Domain: signal sequence #status predicted <SIG>
F:22-36/Domain: propeptide #status predicted <PRO>
F:37-477/Product: plasminogen activator alpha-2 #status predicted <PLA>
F:42-79/Domain: fibronectin type I repeat homology <IFA>
F:87-120/Domain: EGF homology <EGF>
F:128-209/Domain: kringle homology <KRG>
F:226-471/Domain: trypsin homology <TRY>
F:42-72,70-79,87-98,92-109,111-120,128-209,149-191,180-204,214-345,257-273,265-334,359-4
F:185,398/Binding site: carbonylhydrolase (Asn) (covalent) #status predicted
F:225-226/cleavage site: His-Ser (plasmin) #status predicted
F:272,321,428/Active site: His, Asp, Ser #status predicted

Query Match 23.2%; Score 731; DB 2; Length 477;
Best Local Similarity 35.3%; Pred. No. 3.5e-38;
Matches 165; Conservative 74; Mismatches 170; Indels 58; Gaps 16;

QY 115 CKDNPGRGQCLITQSPPIYR-----CVCKHPYTPGSCSQVVPV--CRPNPCQNG 162
DB 42 CRDE---KTQMIYQQESMLRPEVRSKRVEHCRCDR---GLAQCHTVPVKSCSELRCFNG 95
QY 163 ATCSRHKRSKFTACPDQPKGKFCIGSD-DCYVGDGYSYRGKMTNVTQVHACLYWNH 221
DB 96 GTCWQAASFDFVCCPKGTGKQCEVDTHATCYKQGVYRGVTSSTSGAQCINWNSN 155
QY 222 LLLQENTYMFMDAETHGIGEHNFRCNPDADKPWCIFKVTNDKVKWEYCDVSACSQAQV 281
DB 156 LLTRRTYNGRRSDAITLGLGNHNYCRPNPNNSKPCVY-IKASKFILEFCSPVCSKA-- 212
QY 282 AYPESPTSTKLPFGDSCGKTEIAERKIKRIYGGFKSTAGKHFWQASLQSLPLTISM 341
DB 213 -----TCGLRKYPQLHST-GGLFTDITSHPWQAIFAQ--NRRSS 251
QY 342 PQGHFCGGLIHPCWLTAAHCTDIK--TRHLKVLVGDQDLKKEEFHQSFRVEKIFKYS 399
DB 252 GERFLCGGILLISSCWLTAAHCFQERYPPQHLRVVLGRVTVRKVPKGEQTFEVEKICIVHE 311

```
QY 400 HYNREDEIPNDIALLKLEVDGHCALSKYKVTCLPDGS--FPSGSECHISGVTET 457
Db 312 EFD--DNTYNDIALLQKSGSQCAQSDSVRAICLPEANLQLPDWTCELSGYGKHS 369
QY 458 GKG--SRQLDAKVKIANTLNSROLYDHIMDDSMICAGNLQK-----PG-QDTCQDGG 510
Db 370 SSPFYSEQLKEGHVRLYPSRCKSKFLNKTNTNNMLCAGDTRSGBIYENVAHQDQDGG 429
QY 511 GPITCEKDGYYVYVGVSMGLECGKR--PGVYTVQTKFLNWKATIK 555
Db 430 GPLVCMNDNMTLLGIISWVGCGEKDIPGVYTKVNYLWGRDNR 476

RESULT 11
JS0599
t-plasminogen activator (EC 3.4.21.68) beta precursor - common vampire bat
N:Alternate names: tissue plasminogen activator
C:Species: Desmodus rotundus (common vampire bat)
C>Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 16-Jul-1999
C:Accession: JS0599
R:Kraetzschmar, J.; Haendler, B.; Langer, G.; Boidol, W.; Bringmann, P.; Alagon, A.; Don
Gene 105, 229-237, 1991
A:Title: The plasminogen activator family from the salivary gland of the vampire bat Des
A:Reference number: JS0597; MUID:92039036; PMID:1937019
A:Accession: JS0599
A:Molecule type: mRNA
A:Residues: 1-431 <KRA>
A:Cross-references: GB:M63989; NID:G166076; PID:AAA1594.1; PID:G166077
C:Superfamily: tissue plasminogen activator; EGF homology; fibronectin type I repeat hom
C:Keywords: fibrinolysis; glycoprotein; hydrolase; kringle; serine proteinase
F:1-21/Domain: signal sequence #status predicted <SIG>
F:22-36/Domain: propeptide #status predicted <PRO>
F:37-43/Product: plasminogen activator beta #status predicted <PLA>
F:41-74/Domain: EGF homology <EGF>
F:82-163/Domain: kringle homology <KR>
F:180-425/Domain: trypsin homology <TRY>
F:41-52,46-63,65-74,82-163,103-145,134-158,168-299,211-227,219-288,313-388/Disulfide bon
F:139,352/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:179-180/cleavage site: His-Ser (plasmin) #status predicted
F:226,275,382/Active site: His, Asp, Ser #status predicted
F:345-361,378-406/Disulfide bonds: #status predicted

Query Match 22.8%; Score 720.5; DB 2; Length 431;
Best Local Similarity 36.0%; Pred. No. 1.4e-37;
Matches 157; Conservative 73; Mismatches 159; Indels 47; Gaps 13;

QY 134 YRCVKHPYTGPSGVVPCRPNCPONGATCSRHKRSKFTACPDQPKGKCEIGSD- 192
Db 28 YRQLARGRAYGGCSEL-----RCNFGTCWAASFDFVCCQPKGYTGKQCEVDTHA 80
QY 193 DCYVGDGYSYRGKMRNTVNHACLYWNHLLQENYMFMEDEATHGIGEHNFGRNPDAD 252
Db 81 TCYKDGQVYTRGTWSTSEGAQCINWNSNLLTRTYNGRRSDAITLGLGNHNYCRNPDN 140
QY 253 EKPWCFTKVNWKVWEYCDVSAGSAQDVAYPESPTSTKLPFGPSCGKTEIAERKIK 312
Db 141 SKPWCYV-IKASKETILEFCSPVPCSKA-----TCGLRYKKEPOLH 179
QY 313 RIYGFKGTAGHPMQASLQSLPLTISMPQGHFCGGALHPCVWLTAACHCTDIK--TRH 370
Db 180 ST-GGLFTDITSHPPQAAIPAQ--NRRSSGERFLCGGLISSCVWLTAACHCFQERYPPQH 236
QY 371 LKVLGDQDLKKEBHEHQSFRVEKIFKYSHYNERDEIPNDIALLKLPVDGHCALSKY 430
Db 237 LRVVLGRTYRVKPGKEEQTFEVEKCIIEEFD--DDTYNDIALLQLKSGSPQCAQSDS 294
QY 431 VKTVCCLPDGS--FPSGSECHISGVMGTETGKG--SRQLDAKVKIANTLNSROLYDHIM 486
Db 295 VRAICLPEANLQLPDWTCELSGYGKHSFPFSEQLKEGHVRLYPSRCKSKFLNKT 354
QY 487 IDDSMICAGNLQK-----PG-QDTCQDGGGGLTCEKDTGYVYVGVSMGLECGKR--PGV 539
Db 487 IDDSMICAGNLQK-----PG-QDTCQDGGGGLTCEKDTGYVYVGVSMGLECGKR--PGV 539
```

```
Db 355 VTNNMLCAGDTRSGEYIPNVHDAQQDGGGLVCMNDNMTLLGIISWVGCGEKDIPGV 414
QY 540 YTVTKFLNWKATIK 555
Db 415 YTKVNYLWGRDNR 430

RESULT 12
S45281
coagulation factor XIa (EC 3.4.21.38) precursor - bovine (fragment)
N:Alternate names: Hageman factor (activated)
C:Species: Bos primigenius taurus (cattle)
C>Date: 10-Apr-1995 #sequence_revision 22-Apr-1995 #text_change 21-Jan-2000
C:Accession: S45281; A61329
R:Shibuya, Y.; Semba, U.; Okabe, H.; Kambara, T.; Yamamoto, T.
Biochim. Biophys. Acta 1206, 63-70, 1994
A:Title: Primary structure of bovine Hageman factor (blood coagulation factor XII): com
A:Reference number: S45281; MUID:94242782; PMID:8186251
A:Accession: S45281
A:Molecule type: mRNA
A:Residues: 1-593 <SHI>
A:Cross-references: GB:S70164
A:Note: the authors translated the codon GAG for residue 23 as Val, GAG for residue 70;
s Pro, CTC for residue 203 as Phe, GTG for residue 247 as Leu, CCG for residue 286 as G;
is, and ATC for residue 505 as Leu
R:Fujikawa, K.; Walsh, K.A.; Davie, E.W.
Biochemistry 16, 2270-2278, 1977
A:Title: Isolation and characterization of bovine factor XII (Hageman factor).
A:Reference number: A61329; MUID:77181112; PMID:861210
A:Accession: A61329
A:Molecule type: protein
A:Residues: 10-16,'X',18-19,525-550 <FUJ>
C:Superfamily: coagulation factor XII; EGF homology; fibronectin type I repeat homology
C:Keywords: blood coagulation; fibrinolysis; glycoprotein; hydrolase; monomer; plasma;
F:37-78/Domain: fibronectin type II repeat homology <IF2>
F:78-120/Domain: EGF homology <EGF>
F:125-160/Domain: fibronectin type I repeat homology <FB1>
F:207-287/Domain: kringle homology <KR>
F:350-587/Domain: trypsin homology <TRY>
F:541/Active site: Ser #status predicted

Query Match 22.4%; Score 706.5; DB 2; Length 593;
Best Local Similarity 33.0%; Pred. No. 1.4e-36;
Matches 191; Conservative 69; Mismatches 183; Indels 135; Gaps 31;

QY 56 SSTLTHAENPDWYY---TEDQADPC-QNPCEHGGDCLVHGSTFTCSCLAPFGSNKCKV 111
Db 64 ATTFNFKDQKWAYCLEPKVKVCKSKENPCQKGGTCVNMNPDGPRCICADHFTGKHKOK- 122
QY 112 QNTCKNCGRCQCLITQSPVYR-----CVCKHP--YTGFSQSQWV 151
Db 123 -----EKCF---EPOFFRFFHENEIWHLEPAGVYVCKCKGPNACKPLASQ-- 166
QY 152 PVCPRNPONGATCSR---HKRSKFTACPDQPKGKCEIG-SDDCY--VGDGYSYRGK 205
Db 167 -VCRINPCLNGSDCLQAEGRH-----LCRCAPSAGRLCDVDLKASCYDDRDRLGSLYRGM 220
QY 206 MERTVNHACLYWNHLLQENYMFMEDEATHGIGEHNFGRNPDADSKPCFIKVTNDK 265
Db 221 AGTTLGAPCQSWASEAY---WNVTAEVLNWLGDHAFGRNPDNDTRPWCPI-WKGR 276
QY 266 VKWEYCDVSAC--SAQDVAYPEEP-----TEPSTKLPGFDS--CKTEIAE----- 308
Db 277 LSNVYCRCLAPCAAAAGHEHFPPLPSALQKPESTTQTPSLTSGWCSPTPLASGGPGGC 336
QY 309 -----RK-----IKRIVGGFKSTAGKHPQASLOSPLTISMPQGHFCGGALHPCVWLTA 360
Db 337 GORLRKWLSSLRVVGVGLVALFGAHPYTAALYW-----DQHFCAGLIAPCVWLTA 387
QY 361 AHCTDIK--TRHLKVVVLGDQLKKEBHEHQS-----FRVEKIFK---YSHYNER 404
Db 388 AHCLQNEPAPKELTVVLG-----QDRHNSQCEQCTQTLAVRDYRLHEAFSPITYQH---- 437
```

QY 405 DEIPNDIALKL-KPVDGCHCALESKYVTKVLCIPD-GSFPSSGSE---CHISGWG-VTETG 458
DB 438 -----DLALVRLQESADGCCAHPSPFPVQVCLPSTFAARPAESAAVCEVAGNWHQFEGG 491
QY 459 KGSROLLDAKVLANTLCNSRQLYDHMDISMCAGNLQKPGQDTCQGSQGGPLTCE-- 516
DB 492 EYSSFLQEAQVPLIDPQRCSPADVHGAAFTQGMCLCAGFLE--GGTACQDQSGGFLVCEDE 550
QY 517 -KDGYYVYGVISWGLECGKR--PGVYTVQTKFLNWK 551
DB 551 TPERQLILRGVSGWGCGNRLKPGVYTDVANYLAWIR 588
RESULT 13
J50597
t-plasminogen activator (EC 3.4.21.68) alpha-1 precursor - common vampire bat
N:Alternate names: tissue plasminogen activator
C:Species: Desmodus rotundus (common vampire bat)
C>Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 16-Jul-1999
C:Accession: J50597
R:Kratzschmar, J.; Haendler, B.; Langer, G.; Boidol, W.; Bringmann, P.; Alagon, A.; Don
Gene 105, 229-237, 1991
A:Title: The plasminogen activator family from the salivary gland of the vampire bat Des
A:Reference number: J50597; MUID:92039036; PMID:1937019
A:Accession: J50597
A:Molecule type: mRNA
A:Residues: 1-477 <KRA>
A:Cross-references: GB:M63987; NID:g166070; PIDN:AAA31591.1; PID:g166071
C:Superfamily: tissue plasminogen activator; EGF homology; fibronectin type I repeat hom
C:Keywords: fibrinolysis; glycoprotein; hydrolase; kringle; serine proteinase
F:1-21/Domain: signal sequence #status predicted <SIG>
F:22-36/Domain: propeptide #status predicted <PRO>
F:37-477/Product: plasminogen activator alpha-1 #status predicted <PLA>
F:42-79/Domain: fibronectin type I repeat homology <1FA>
F:87-120/Domain: EGF homology <EGF>
F:128-209/Domain: kringle homology <KRG>
F:226-471/Domain: trypsin homology <TRY>
F:42-72,70-79,87-98,92-103,111-120,128-209,149-191,180-204,214-345,257-273,265-334,359-4
F:153,398/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:225-226/Cleavage site: His-Ser (plasmin) #status predicted
F:272,321,428/Active site: His, Asp, Ser #status predicted
Query Match 22.3%; Score 704.5; DB 2; Length 477;
Best Local Similarity 35.9%; Pred. No. 1.5e-36;
Matches 161; Conservative 64; Mismatches 160; Indels 63; Gaps 15;
QY 120 CGRQGLITQSPPYRVCVKCHPTGSCSQVVPV--CRPNPCQNGATCSRHKRSKFTCA 177
DB 72 CDRGQA-----RC-----HTVPVNSCGEPRCFNGTQWQAVYFSDVFCQ 110
QY 178 CPDQPKGKCEIGS-DDCVGDSYSGYRGKQNTVQHACLWNSHLLLOENYMFEMDAE 236
DB 111 CPAGYTGKRCVDTTRATCYEGQGVYRGTWSTAESRVCINWNSLLTTRTYNGRMPDAF 170
QY 237 THGIGHNFCRNPDADEKPCWCFIKVTNDKVKWEYCDVGSACSQDVAYPEESPTETKLP 296
DB 171 NLGLGNHYCRNPNGAPKWCIV--IKAGFTSESQVPCSKA----- 212
QY 297 GFDSCGKTEIAERKIKRIYGGPKSTAGKHPWQASLQSSPLTISMPOGHFGGALIHPCW 356
DB 213 --TCGLRKYKBPQLHST--GGLFTDITSHPWQAALFAQ--NRRSSGERFLCGGILLSCW 266
QY 357 VLTAHAC--TDIKTRHLKVVLGDQDLKKEEFHQSFVEKIKYSHYNERDEIPNDIAL 414
DB 267 VJTAACCFQESVLPDQGLKVLGRTYRKVKEEQTFFVKYIVHKEFD--DDTYNDIAL 324
QY 415 LKLKPDVGHCALESKYVTKVLCIPDGS--FPSSGECHISGWVTFETGK--SSQLLDKVK 470
DB 325 LQLKSDSPQCAQESDSVRAICLPEANLQLPDWTECELSGYGKHSSPYSQLKXGHRV 384
QY 471 LTANTLCNSRQLYDHMDISMCAGNLQK-----PG-QDPCQDSSGGLTCEKDGYYVYG 525
DB 385 LYPSSRCAPKFLFNKTVTNMLCAGDTRSGEYIPNVHDCQDQSGGFLVCMNDNENTLLG 444

QY 526 IVSWGLECGKR--PGVYTVQTKFLNWK 551
DB 445 IISWGVCGEKDVPQGVYTKVNYLGNIR 472

RESULT 14

S18932

u-plasminogen activator (EC 3.4.21.73) precursor - rat
N:Alternate names: plasminogen activator, urokinase-type; urinary plasminogen activator
C:Species: Rattus norvegicus (Norway rat)
C>Date: 18-Oct-1989 #sequence_revision 10-Feb-1995 #text_change 18-Jun-1999
C:Accession: S24604; I60186; I53472; S18932
R:Rabbani, S.A.
submitted to the EMBL Data Library, April 1992
A:Reference number: S24604
A:Accession: S24604

A:Molecule type: mRNA
A:Residues: 1-15, 'H', 17-23, 'G', 25-331, 'N', 333-432 <RAB>
A:Cross-references: EMBL:X65651; NID:g57456; PIDN:CAA46601.1; PID:g57457
R:Experimental source: tissue kidney
R:Handerson, B.R.; Tansey, W.P.; Phillips, S.M.; Ramshaw, I.A.; Kefford, R.F.
Cancer Res. 52, 2489-2496, 1992
A:Title: transcriptional and posttranscriptional activation of urokinase plasminogen act
A:Reference number: I60186; MUID:92233409; PMID:1568219
A:Accession: I60186
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-432 <RES>
A:Cross-references: EMBL:X63434; NID:g57465; PIDN:CAA45028.1; PID:g57466
R:Experimental source: strain Fischer 344; tissue mammary
R:Ragno, P.; Cassano, S.; Degen, J.; Kessler, C.; Blasi, F.; Rossi, G.
FEBS Lett. 306, 193-198, 1992
A:Title: The receptor for the plasminogen activator of urokinase type is up-regulated in
A:Reference number: I53472; MUID:92339549; PMID:1321734
A:Accession: I53472
A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA
A:Residues: 31-62 <RE2>
A:Cross-references: EMBL:X66907; NID:g396200; PIDN:CAA47356.1; PID:g938279
C:Genetics:
A:Gene: uPA
C:Superfamily: urokinase-type plasminogen activator; EGF homology; kringle homology; try
C:Keywords: glycoprotein; heterodimer; hydrolase; kringle; serine proteinase
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-177/Product: urokinase-type plasminogen activator chain A #status predicted <ACH>
F:31-62/Domain: EGF homology <EGF>
F:70-151/Domain: kringle homology <KRG>
F:179-432/Product: urokinase-type plasminogen activator chain B #status predicted <BCH>
F:179-420/Domain: trypsin homology <TRY>
F:168-300,210-226,218-289,314-383,346-362,373-401/Diulfide bonds: #status predicted
F:225,276,377/Active site: His, Asp, Ser #status predicted

Query Match 22.3%; Score 704; DB 1; Length 432;
Best Local Similarity 37.9%; Pred. No. 1.5e-36;
Matches 157; Conservative 64; Mismatches 163; Indels 30; Gaps 14;

QY 159 CQNGATCSRHKRSKFTACAPDQPKGKCEIG-SDDCVGDSYSGYRGKQNTVQHACL 216
DB 33 CQNGCVSVTRYFSSRRSCPKFKGEHCEIDTSTCYHGNGQSYRGKANTDKRPCL 92
QY 217 YWNSHLLQBNYMFEMDAETHGIBENFCRNPDADEKPCWCFIKVTNDKVKWEYCDVSAC 276
DB 93 ANNSPAVLQQTYNNAHRSALSLGLGKHNYCRPNQRPWCYVQIGLKQF-----VQEC 146
QY 277 SAQDVAYPEESPTETKLP--PGFDSGCKTEIAERKIKRYGGPKSTAGKHPWQASLQSS 334
DB 147 MVQDCSLSK----KPSSTVDQQGF--QCGQVALRPR--FKLVGGEFTVVENQPFALY-- 197
QY 335 LPLTISMPOGHFGGALIHPCWVLTAAHC--TDIKTRHLKVVLGDQDLKKEEFHQSFV 392
DB 198 LKNKGSGSPSPKCGSLISPCWASATHCFVNPQKKEEYVYVYLGQSKRNSYNPGENMKFEY 257

Search completed: May 24, 2004, 09:49:41
Job time : 17.5 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model
Run on: May 24, 2004, 09:40:10 ; Search time 11.5 Seconds
(without alignments)
2535.590 Million cell updates/sec

Title: US-09-912-559-3
Perfect score: 3154
Sequence: 1 MFARMSDLHVLLMALVGT.....TOVTKPLNLIKATIKSESGF 560

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues
Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	832.5	26.4	655	1 HGFA_HUMAN	Q04756 homo sapien
2	829	26.3	653	1 HGFA_MOUSE	Q95098 mus musculus
3	773.5	24.5	603	1 FA12_CAVPO	Q04962 cavia porce
4	766.5	24.3	562	1 TPA_HUMAN	P00750 homo sapien
5	757.5	24.0	559	1 TPA_RAT	P19637 rattus norv
6	753.5	23.9	559	1 TPA_MOUSE	P11214 mus musculus
7	741	23.5	615	1 FA12_HUMAN	P00748 homo sapien
8	734.5	23.3	566	1 TPA_BOVIN	Q28198 bos taurus
9	731	23.2	477	1 URT2_DESRO	P15638 desmodus ro
10	720.5	22.8	431	1 URTB_DESRO	P98121 desmodus ro
11	706.5	22.4	593	1 FA12_BOVIN	P98140 bos taurus
12	704.5	22.3	477	1 URT1_DESRO	P98119 desmodus ro
13	704	22.3	432	1 UROK_RAT	P29598 rattus norv
14	686	21.8	442	1 UROK_PIG	P04185 sus scrofa
15	683.5	21.7	433	1 UROK_BOVIN	Q05589 bos taurus
16	680	21.6	433	1 UROK_MOUSE	P06869 mus musculus
17	660	20.9	433	1 UROK_PAPCY	P16227 papio cynoc
18	660	20.9	812	1 PLMN_BOVIN	P06868 bos taurus
19	658.5	20.9	431	1 UROK_HUMAN	P00749 homo sapien
20	650.5	20.6	434	1 UROK_CHICK	P15120 gallus gall
21	643.5	20.4	394	1 URTG_DESRO	P49150 desmodus ro
22	642.5	20.4	810	1 PLMN_HUMAN	P00747 homo sapien
23	635.5	20.1	810	1 PLMN_MACMU	P12545 macaca mula
24	607	19.2	790	1 PLMN_PIG	P06867 sus scrofa
25	603.5	19.1	812	1 PLMN_MOUSE	P20918 mus musculus
26	592.5	18.8	4548	1 APOA_HUMAN	P08519 homo sapien
27	578.5	18.3	1420	1 APOA_MACMU	P14417 macaca mula
28	571.5	18.1	1420	1 PLMN_MACMU	Q29485 erinaceus e
29	558	17.7	338	1 PLMN_ERIEU	P80010 equus cabal
30	545	17.3	333	1 PLMN_HORSE	P80009 canis famil
31	535.5	17.0	343	1 PLMN_CANFA	P81286 ovis aries
32	467.5	14.8	475	1 FA10_CHICK	P25155 gallus gall
33	462.5	14.7	761	1 NETR_MOUSE	O08762 mus musculus

ALIGNMENTS

RESULT 1

ID	HGFA_HUMAN	STANDARD	PRT	655 AA
AC	Q04756; Q14726			
DT	01-JUN-1994 (Rel. 29, Created)			
DT	01-JUN-1994 (Rel. 29, Last sequence update)			
DT	10-OCT-2003 (Rel. 42, Last annotation update)			
DE	Hepatocyte growth factor activator precursor (EC 3.4.21.-) (HGF activator) (HGFA)			
DE	activator) (HGFA)			
GN	HGFAC			
OS	Homo sapiens (Human)			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
CX	NCBI_TaxID=9606;			
RA	Kitamura N.; Miyazawa K.; Shimomura T.; Kitamura A.; Kondo J.; Morimoto Y.;			
RA	Kitamura N.;			
RT	"Molecular cloning and sequence analysis of the cDNA for a human serine protease responsible for activation of hepatocyte growth factor. Structural similarity of the protease precursor to blood coagulation factor XII."			
RT	coagulation factor XII."			
RT	J. Biol. Chem. 268:10024-10028(1993).			
RL	(2)			
RL	SEQUENCE OF 40-655 FROM N.A.			
RN	Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.			
RP	Zhao S., Odell C.;			
RA	Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.			
RL	Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.			
CC	-!- FUNCTION: Activates hepatocyte growth factor (HGF) by converting it from a single chain to a heterodimeric form.			
CC	-!- SUBUNIT: Dimer of a short chain and a long chain linked by a disulfide bond.			
CC	-!- SUBCELLULAR LOCATION: Secreted as an inactive single-chain precursor and is then activated to a heterodimeric form.			
CC	-!- TISSUE SPECIFICITY: Liver.			
CC	-!- SIMILARITY: Belongs to peptidase family S1.			
CC	-!- SIMILARITY: Contains 2 EGF-like domains.			
CC	-!- SIMILARITY: Contains 1 fibronectin type I domain.			
CC	-!- SIMILARITY: Contains 1 fibronectin type II domain.			
CC	-!- SIMILARITY: Contains 1 kringle domain.			
CC	-!- CAUTION: It is uncertain whether Met-1 is the initiator.			

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@sb-sib.ch).

EMBL; D14012; BAA03113.1; -;
EMBL; Z69923; -; NOT_ANNOTATED_CDS.
PIR; A46688; A46688.
HSSP; P00763; IDPO.
MEROPS; S01.228; -;

```
DR Genew; HGNC:4894; HGFAc.
DR MLM; 604552; .. C:extracellular; TAS.
DR GO; GO:000576; F:serine-type endopeptidase activity; TAS.
DR GO; GO:004252; F:serine-type endopeptidase activity; TAS.
DR GO; GO:0004508; P:proteolysis and peptidolysis; TAS.
DR Inter-Pro; IPR009003; Cys Ser trypsin.
DR Inter-Pro; IPR000742; EGF_2.
DR Inter-Pro; IPR006209; EGF_like.
DR Inter-Pro; IPR000083; Fibronectin.
DR Inter-Pro; IPR000562; FN_Type_I.
DR Inter-Pro; IPR006210; IEGF.
DR Inter-Pro; IPR000001; Kringle.
DR Inter-Pro; IPR001254; Peptidase_S1.
DR Inter-Pro; IPR001314; Peptidase_S1A.
DR Pfam; PF00008; EGF_2.
DR Pfam; PF00039; fn1; 1.
DR Pfam; PF00040; fn2; 1.
DR Pfam; PF00051; kringle; 1.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR PRINTS; PR00013; FNTYPEII.
DR PRINTS; PR00018; KRINGLE.
DR ProDom; PD000935; FN_Type_II; 1.
DR ProDom; PD000395; Kringle; 1.
DR SMART; SM00181; EGF; 2.
DR SMART; SM00058; FN1; 1.
DR SMART; SM00059; FN2; 1.
DR SMART; SM00130; KR; 1.
DR SMART; SM00020; Tryp_Spc; 1.
DR PROSITE; PS00022; EGF_1; 2.
DR PROSITE; PS01185; EGF_2; 1.
DR PROSITE; PS00028; EGF_3; 2.
DR PROSITE; PS01253; FIBRONECTIN_1; 1.
DR PROSITE; PS00023; FIBRONECTIN_2; 1.
DR PROSITE; PS00021; KRINGLE_1; 1.
DR PROSITE; PS00070; KRINGLE_2; 1.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
DR Hydrolase; Glycoprotein; Plasma; Serine protease; Kringle; Signal;
KW EGF-like domain; Repeat; Zymogen.
FT SIGNAL 1 30
FT PROPEP 31 372
FT CHAIN 373 407
FT FT
FT FT
FT FT
FT CHAIN 408 555
FT FT
FT DOMAIN 108 148
FT DOMAIN 160 198
FT DOMAIN 200 240
FT DOMAIN 241 279
FT DOMAIN 286 367
FT DOMAIN 408 555
FT ACT_SITE 447 447
FT ACT_SITE 497 497
FT ACT_SITE 598 598
FT DISULFID 108 133
FT DISULFID 122 148
FT DISULFID 164 175
FT DISULFID 169 186
FT DISULFID 188 197
FT DISULFID 202 230
FT DISULFID 228 237
FT DISULFID 245 256
FT DISULFID 250 267
FT DISULFID 269 278
FT DISULFID 367 367
FT DISULFID 307 349
FT DISULFID 338 362
FT DISULFID 394 521
FT DISULFID 432 448
FT DISULFID 440 510
FT DISULFID 535 604
```

```
FT DISULFID 567 583
FT DISULFID 594 622
FT CARBOHYD 48 48
FT CARBOHYD 290 290
FT CARBOHYD 468 468
FT CARBOHYD 492 492
FT CARBOHYD 546 546
FT CONFLICT 644 644
SQ SEQUENCE 655 AA; 70681 MW; 2CF72F1E1B862ED7 CRC64;

Query Match 26.4%; Score 832.5; DB 1; Length 655;
Best Local Similarity 33.2%; Pred. No. 2.5e-51;
Matches 192; Conservative 87; Mismatches 216; Indels 83; Gaps 21;

QY 31 ESLDPDWTPQDYSDYEDYNQENSTLTHANPDW-----YTTDOA----- 74
DQ 99 QALTEDGRPCRPFRYGRMLHACTSEGSARHK---WCATTHYDRDRAWGYCVERATPPP 155
QY 75 -----DPCQNPCEHGDC--LVHGSTFTCSCLAPFSGNKCQKQVQNTCDNFCGRGQCL 126
DQ 156 GPPAALDPCASGCLNGSGCSNTQDPOSYHCSCPRAFTG-----KD--CGTEKCF 203
QY 127 ITQSPYY-----RCVKHPTGSPSCQVV--PVCRNPCQNGATCSRHK 169
DQ 204 DETRYEVLGGDRWARVRQGHVEQEC---FGRTWCEGTRHTACLSPLCLNGGTCHLIV 260
QY 170 RRSKFTACDPQKQKCEIGSD--CVVGDGYSYRGQNRVTNQHACLYWNSHLLQENY 228
DQ 261 ATGTVACACPPGFAGRLCNIEPDERCFLNGTGYRGVASTSASGLSCLANNSDLLYQELH 320
QY 229 NMFMEDATHGIGBEHFCRNPDADKPCFKVTNDKVKWEYCDVSACSA--QDVAYPEE 286
DQ 321 VDSGAAALLGLGPHAYCRNPNDNERPCYV--VKDSALSWEYCRLEACEBSLTVQLSPDL 379
QY 287 SPTSPSTKLPQFDSGCKTEAERKIK--RIYGGFKSTAGKHPQASLOSPLTISMPOGH 345
DQ 380 LATUPEPASPORQACGRHKKRTFLRPIRIGGSSSLPGSHFWLAAIYG-----DS 430
QY 346 FCGGALIHPCVWLTAHCTDIKTRH--LKVLGQDQDLKKEEFHQSPRVKIFKYSYNE 403
DQ 431 FCAGSLVHTCVVSAACHFSPPRDSVSVVLGQHFNRRTDVTQTGIEKIYPTILYSV 490
QY 404 RDEPHNDIALKLPVDGHCALSKVKVTVCPLD--GSPSPGSECHISGNG-VTETGKG 460
DQ 491 FNPDSH-DLVIRLKKKGDRCATCSQFVQPCILPEPGSTFPAGHKQIAGHJLDENVSG 549
QY 461 -SRQLLDARVLIANTLNSQLYDHMIDDSMICAGNLQKPGQDTCCGDSGGPLTCRKG 519
DQ 550 YSSSLREALVPLVADHKCSSPEVYGADISPNMLCAGYFDCK-SDACQDSGGPLACRKG 608
QY 520 TYVTGIVSWGLECK--RPGVYQVTKFLNWKATIK 555
DQ 609 VAYLYGIISWGDGCGRLHKPGYTRVANYVDNDRIR 646

RESULT 2
HGFA_MOUSE STANDARD; PRT; 553 AA.
AC Q9R098; Q9JUV4;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Hepatocyte growth factor activator precursor (EC 3.4.21.-) (HGF activator) (HGFA).
DE HGFAc.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/c;
RA Itch H., Kataoka H., Kono H.;
```

RT "Mouse hepatocyte growth factor activator.":
 RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
 RN (2)
 RX SEQUENCE FROM N.A.
 RA MEDLINE-2126753; PubMed-11032833;
 RA van Adelsberg J.S., Sehgal S., Kukes A., Brady C., Barasch J.,
 RA Yang J., Huan Y.;
 RT "Activation of hepatocyte growth factor (HGF) by endogenous HGF
 RT activator is required for metanephric kidney morphogenesis in
 RT vitro".
 RT J. Biol. Chem. 276:15099-15106(2001).
 CC -!- FUNCTION: Activates hepatocyte growth factor (HGF) by converting
 CC it from a single chain to a heterodimeric form (By similarity).
 CC -!- SUBUNIT: Dimer of a short chain and a long chain linked by a
 CC disulfide bond (By similarity).
 CC -!- SUBCELLULAR LOCATION: Secreted as an inactive single-chain
 CC precursor and is then activated to a heterodimeric form (By
 CC similarity).
 CC -!- SIMILARITY: Belongs to peptidase family S1.
 CC -!- SIMILARITY: Contains 2 EGF-like domains.
 CC -!- SIMILARITY: Contains 1 fibronectin type I domain.
 CC -!- SIMILARITY: Contains 1 fibronectin type II domain.
 CC -!- SIMILARITY: Contains 1 kringle domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AF099017; AAF02489.1; -;
 CC EMBL; AF224724; AAF34712.1; -;
 CC HSSP; P00763; IDPO.
 CC MEROPS; S01.228; -;
 CC MGD; MGI:1859281; Hgfac.
 CC InterPro; IPR009003; Cys_ser_trypsin.
 CC InterPro; IPR000742; EGF_2.
 CC InterPro; IPR006209; EGF_Like.
 CC InterPro; IPR000083; Fibrnctnl.
 CC InterPro; IPR000562; FN_Type_II.
 CC InterPro; IPR006210; IEGF.
 CC InterPro; IPR000001; Kringle.
 CC InterPro; IPR001254; Peptidase_S1.
 CC InterPro; IPR001314; Peptidase_S1A.
 CC Pfam; PF00008; EGF; 2.
 CC Pfam; PF00039; fn1; 1.
 CC Pfam; PF00040; fn2; 1.
 CC Pfam; PF00051; kringle; 1.
 CC Pfam; PF00089; trypsin; 1.
 CC PRINTS; PR00722; CHYMOTRYPSIN.
 CC PRINTS; PR00013; ENTYPEII.
 CC PRINTS; PR00018; KRINGLE.
 CC ProDom; PD000995; FN_Type_II; 1.
 CC ProDom; PD000395; Kringle; 1.
 CC SMART; SM00181; EGF; 2.
 CC SMART; SM00059; FN2; 1.
 CC SMART; SM00130; KE; 1.
 CC SMART; SM00020; Tryp_Spc; 1.
 CC PROSITE; PS00022; EGF_1; 2.
 CC PROSITE; PS01186; EGF_2; 1.
 CC PROSITE; PS00026; EGF_3; 2.
 CC PROSITE; PS01233; FIBRONECTIN_1; 1.
 CC PROSITE; PS00023; FIBRONECTIN_2; 1.
 CC PROSITE; PS00021; KRINGLE_1; 1.
 CC PROSITE; PS00070; KRINGLE_2; 1.
 CC PROSITE; PS02400; TRYPsin_DOM; 1.
 CC PROSITE; PS00134; TRYPsin_HIS; 1.
 CC PROSITE; PS00135; TRYPsin_SER; 1.
 CC KW Hydrolase; Glycoprotein; Plasma; Serine protease; Kringle; Signal;
 CC EGF-like domain; Repeat; Zymogen.
 CC SIGNAL 1 29 BY SIMILARITY.

FT PROPEP 30 369 CLEAVED IN ACTIVE FORM (BY SIMILARITY).
 FT CHAIN 370 405 HEPATOCYTE GROWTH FACTOR ACTIVATOR SHORT
 FT CHAIN 406 653 HEPATOCYTE GROWTH FACTOR ACTIVATOR LONG
 FT DOMAIN 105 145 FIBRONECTIN TYPE-II.
 FT DOMAIN 157 195 EGF-LIKE 1.
 FT DOMAIN 197 237 FIBRONECTIN TYPE-I.
 FT DOMAIN 238 276 EGF-LIKE 2.
 FT DOMAIN 283 364 KRINGLE.
 FT DOMAIN 406 653 SERINE PROTEASE.
 FT ACT_SITE 445 445 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 495 495 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 596 596 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT DISULFID 109 130 BY SIMILARITY.
 FT DISULFID 119 145 BY SIMILARITY.
 FT DISULFID 161 172 BY SIMILARITY.
 FT DISULFID 166 183 BY SIMILARITY.
 FT DISULFID 185 194 BY SIMILARITY.
 FT DISULFID 199 227 BY SIMILARITY.
 FT DISULFID 225 234 BY SIMILARITY.
 FT DISULFID 242 253 BY SIMILARITY.
 FT DISULFID 247 264 BY SIMILARITY.
 FT DISULFID 266 275 BY SIMILARITY.
 FT DISULFID 283 364 BY SIMILARITY.
 FT DISULFID 304 346 BY SIMILARITY.
 FT DISULFID 335 359 BY SIMILARITY.
 FT DISULFID 392 519 INTERCHAIN (BY SIMILARITY).
 FT DISULFID 430 446 BY SIMILARITY.
 FT DISULFID 438 508 BY SIMILARITY.
 FT DISULFID 533 602 BY SIMILARITY.
 FT DISULFID 565 581 BY SIMILARITY.
 FT DISULFID 592 620 BY SIMILARITY.
 FT CARBOHYD 39 39 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 47 47 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 63 63 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 287 287 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 466 466 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 544 544 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CONFLICT 164 164 G -> W (IN REF. 2).
 SQ SEQUENCE 653 AA; 70567 MW; 88B4B20255DF7FDC CRC64;

Query Match 26.3%; Score 829; DB 1; Length 653;
 Best Local Similarity 34.4%; Pred. No. 4.4e-51;
 Matches 190; Conservative 79; Mismatches 187; Indels 96; Gaps 22;

Qy 60 THAENPD--WYTEDQA-----DPCQNPCEHGGDCLV---HGSTFTCSCLAFFSG 105
 Db 133 THNYDRDRAWGYCAEVTLPVEGPAILDPCASGCLNGGTCSSTHDSGS-YHSCSLAFTG 191
 Qy 106 NKQCKVQNTCKD-----NPGCRGQCLITQSPPYRYRCVKHPYTG 144
 Db 192 KDCG--TEKCFDETREYFEVGDHARVSEGHVEQCG---CMGQA---RCDTHH--- 239
 Qy 145 PSCSQVVPVCRPNPCONGATCSRHKRSKFTACPDQFKGFCFI-GSDDCYVDGYSYR 203
 Db 240 -----TACLSSPCLNGGTCHLIVGTGTSVCTCPILGYAGRCFNIVPTHECFLNGTEYR 292
 Qy 204 GKQNRVTNQHACLYWNSHLLLOENWFMEDASTHGIGEHNFGRNPDADKPCWCFIKVTN 263
 Db 293 GVASTAASGLSCLAWNSDLLYQELHVDVSAVAALLGLGPHAYCRNFDKDERPWCYV-VKD 351
 Qy 264 DKVKEVCDVSACSADQ-----VAYPESPTERTKLPGFDSCKGTEIAERKIK- 312
 Db 352 NALSWEYCRUTACESLARVHSQTPETLAALPESAPAVRPT-----CGKRHKRKTFLRP 404
 Qy 313 RIYGGFKSTAGKHPQASLSQSLPLTISMPQGHFCGALHPCWLVLTAAHC--TDIKTRH 370
 Db 405 RIIGSSSLPGSHPLAAIYG-----NSFCAGSLVHTCWVVAHCFANSPPRDS 455
 Qy 371 LKVLGQDQLKKEEFHQSFREKIPKYSHYNERDEIPHN-DIALKLKLPVDGHCALSK 429
 Db 456 ITVLGQHFFNRRTDVTQTGIEKYVPYTLYSVEN--PNNHDLVLRLKKKGRCARVSRQ 513

activator at 2.4-A resolution.";
 Biochemistry 31:270-279(1992).
 STRUCTURE BY NMR OF KRINGLE 2.
 MEDLINE=90122799; PubMed=2558718;
 Byson I.-J.L., Kelley R.F., Llinas M.;
 "1H NMR structural characterization of a recombinant kringle 2 domain
 from human tissue-type plasminogen activator.";
 Biochemistry 28:9350-9360(1989).
 STRUCTURE BY NMR OF KRINGLE 2.
 MEDLINE=91200042; PubMed=1901789;
 Byson I.-J.L., Kelley R.F., Llinas M.;
 "Kringle-2 domain of the tissue-type plasminogen activator. 1H-NMR
 assignments and secondary structure.";
 Eur. J. Biochem. 197:155-165(1991).
 STRUCTURE BY NMR OF KRINGLE 2.
 MEDLINE=92106329; PubMed=1762144;
 Byson I.-J.L., Llinas M.;
 "Solution structure of the tissue-type plasminogen activator kringle
 2 domain complexed to 6-aminohexanoic acid an antifibrinolytic
 drug.";
 J. Mol. Biol. 222:1035-1051(1991).
 [23]

Query Match 24.3%; Score 766.5; DB 1; Length 562;
 Best Local Similarity 35.0%; Pred. No. 9.4e-47;
 Matches 182; Conservative 77; Mismatches 176; Indels 85; Gaps 20;

QY 77 QCPNPGHGGDC--LVHGSTFTCSCLAPFSGNKKQ--KVQNTCKDNP--CGRG----- 123
 DB 86 CSEPRCFNNGTCCQALYFDFVCCPEGFAGKCEITRATCYEDQGISYRGWSTAESG 145
 QY 124 -QCLITOSPYPYRCVCKHYTPGSCSVVPCVRNPQ-----NGATCSRHKRRKFTAC 178
 DB 146 AECTNWS-----SALAQKPYSGR-----RPDAIRLGIGNHNYCRNPDNRDPSKWCYV 192
 QY 179 PDQFK-GKF-----CEIGSDCYGVDGYSYRGKMRVNOHACLYNNSHLLLOENY 228
 DB 193 ---FRAGKYSSEFCSTACSEGNDCYFGNSAARGTHSLTESGASCLPWNMILICKYV 249
 QY 229 NMFMEDAETHGIGEHNFRCNPDADEKPCWFIKVTNDKYKWEYCDVSACSAQDVAYPEESP 288
 DB 250 TAQNPQAQALGLGKXNYCRNPDGDAKPCWCHV-LKNRLTWECYCDVPSCS----- 297
 QY 289 TEPSTKLPFGSDCGKTEAERKIKRIYGGFKSTAGKHPQASLOSPLTITSMPOGHFCG 348
 DB 298 -----TCGLRQYSQPF-RIRKGLPADIASHPWQAAIFAK--HRRSPGRFLCG 343
 QY 349 GALIHPCWVLTAAHCTD--IKTRHLKVLVGDQDLKKEEFHEQSPRVEKIFKYSHYNERDE 406
 DB 344 GLIISCHILSNAHCFQERFPFHLTVLGTIVYVPCGEEQKFEVEKYIVHKEFD--DD 401
 QY 407 IPHNDIALLKVPDVGCHCALESKYKVCILP--DGSFPGSGSECHISGWVETGKG--SR 462
 DB 402 TYDNDIALQLKSDSSRCAQESSVVRVTCPLPADLPDWTCELSGYGKHEALSPFYSE 461
 QY 463 QLLDAKVKLIANTLCLNSRQLYDHMDSDMICAGNLQKPG-----QDTCCGDSGGPLTCEK 517
 DB 462 RUKERHRLYPSRSTSHLLNRVTYDNLGAGTRSGGPGQANLHDAQCQDSGGPLVCLN 521
 QY 518 DGTYYVYGVISWGLECGKR--PGVYTVQTKFLNLIKATIK 555
 DB 522 DGRMTLVGIISGLGCGKQDVGVTYKVTNYLDWIRDNMR 561

RESULT 5
 TPA RAT
 ID TPA RAT
 AC PI9637;
 DT 01-FEB-1991 (Rel. 17, Created)
 DT 01-NOV-1991 (Rel. 20, Last sequence update)

MEDLINE=85289338; PubMed=3161893;
 Fisher R., Waller E.K., Grossi G., Thompson D., Tizard R.,
 Schleuning W.-D.;
 "Isolation and characterization of the human tissue-type plasminogen
 activator structural gene including its 5' flanking region.";
 J. Biol. Chem. 260:11223-11230(1985).
 [11]
 SEQUENCE OF 31-562 FROM N.A.
 MEDLINE=91291340; PubMed=1369681;
 Itagaki Y., Yasuda H., Morinaga T., Mitsuda S., Higashio K.;
 "Purification and characterization of tissue plasminogen activator
 secreted by human embryonic lung diploid fibroblasts, IMR-90 cells.";
 Agric. Biol. Chem. 55:1225-1232(1991).
 [12]
 SEQUENCE OF 36-562.
 TISSUE=Melanoma;
 MEDLINE=85000468; PubMed=6433976;
 Pohl G., Kaellstroem M., Bergsdorf N., Wallen P., Joernvall H.;
 "Tissue plasminogen activator: peptide analyses confirm an indirectly
 derived amino acid sequence, identify the active site serine residue,
 establish glycosylation sites, and localize variant differences.";
 Biochemistry 23:3701-3707(1984).
 [13]
 SEQUENCE OF 33-52 AND 311-330.
 TISSUE=Melanoma;
 MEDLINE=83209620; PubMed=6682760;
 Wallen P., Pohl G., Bergsdorf N., Raanby M., Ny T., Joernvall H.;
 "Purification and characterization of a melanoma cell plasminogen
 activator.";
 Eur. J. Biochem. 132:681-686(1983).
 [14]
 STRUCTURE OF CARBOHYDRATES.
 MEDLINE=90092112; PubMed=2513186;
 Pfeiffer G., Schmidt M., Strube K.-H., Geyer R.;
 "Carbohydrate structure of recombinant human uterine tissue
 plasminogen activator expressed in mouse epithelial cells.";
 Eur. J. Biochem. 186:273-286(1989).
 [15]
 CARBOHYDRATE-LINKAGE SITE THR-96.
 MEDLINE=91159408; PubMed=1900431;
 Harris R.J., Leonard C.K., Guzzetta A.W., Spellman M.W.;
 "Tissue plasminogen activator has an O-linked fucose attached to
 threonine-61 in the epidermal growth factor domain.";
 Biochemistry 30:2311-2314(1991).
 [16]
 DISULFIDE BONDS IN KRINGLE 2.
 MEDLINE=91244765; PubMed=1645336;
 Vlahos C.J., Wilhelm O.G., Hassell T., Jaskunas S.R., Bang N.U.;
 "Disulfide pairing of the recombinant kringle-2 domain of tissue
 plasminogen activator produced in Escherichia coli.";
 J. Biol. Chem. 266:10070-10072(1991).
 [17]
 X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS) OF CATALYTIC DOMAIN.
 MEDLINE=96200985; PubMed=8613982;
 Lamba D., Bauer M., Huber R., Fischer S., Rudolph R., Kohnert U.,
 Bode W.;
 "The 2.3 A crystal structure of the catalytic domain of recombinant
 two-chain human tissue-type plasminogen activator.";
 J. Mol. Biol. 258:117-135(1996).
 [18]
 X-RAY CRYSTALLOGRAPHY (3.1 ANGSTROMS) OF CATALYTIC DOMAIN.
 MEDLINE=9749126; PubMed=9305622;
 Renatus M., Eng R.A., Stubbs M.T., Huber R., Fischer S., Kohnert U.,
 Bode W.;
 "Lyseine 156 promotes the anomalous proenzyme activity of tPA: X-ray
 crystal structure of single-chain human tPA.";
 EMBO J. 16:4797-4805(1997).
 [19]
 X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS) OF KRINGLE 2.
 MEDLINE=92118803; PubMed=1310033;
 de Vos A., Ultsch M.H., Kelley R.F., Padmanabhan K., Tulinsky A.,
 Westbrook M.L., Kosiakof A.A.;
 "Crystal structure of the kringle 2 domain of tissue plasminogen

QY 96 TSCCLAPFSGNCKO-KVQNTCKNDNPGCGOCLITQSPYYRCVCKHYTPGSC----- 147
Db 104 VQCPDGFVGRKCDITDTRATCFE---GQ3---IT-----YRGTWSTAENGAECINWNSSA 152
QY 148 --SOVVPVCRPNPCGATCSHKRSKFTACPDQ-----FK-GKP-----C 187
Db 153 LQCKPYSARRPNAIKLG--LGNHN-----YCNPDPRVKNPCYVFKAGKVTTFECSTPAC 205
QY 188 EIG-SDDCVVGQYSGYRGKONETVNOHACLYWNSHLLLOENYNNMFMDAETHGIGEHFNC 246
Db 206 PKGPTEDCVVGKGVYRGVTHSTFTSKASCLPWNMILIGKTYTAWRANSOALGLGRHNYC 265
QY 247 RNPDADEKPCWCFIKVNDKWEYCDVSACSAQDVAYPEESTPSTKLPGFDSGCGTEI 306
Db 266 RNPDDGAKPCHV-NKDKLWYCDMSPCS-----TCGLRQY 302
QY 307 ARKIKRIYGGFKSTAGKHPWQASLOSLPLTISWPGHFGGALIHPCWVLRAHC--T 364
Db 303 KQOPF-RIKGGLFTDITSHPWQAAI--FVYKRSFGRFLCGGLVLSICWVLSAHCVE 359
QY 365 DIKTRHLKVLGQDLKKEEFHQFVEKIPKYSHYNERDEIPHNDIALKLPVDPGHC 424
Db 360 RPPPHLKVGLGRTYRVGVEEQFEIEKIVHKEFD--DDTYNDIALQLRSDSSQC 417
QY 425 ALESKYVTVCLPDG--SPPSGSECHISGWGVTETGKG--SRQLLDKVKLIANTLCNSR 480
Db 418 AOESSVSGTACLPDQVQVDFDTECELSGYGKHEASPPFSDRLKEARVLPSPSRCTSQ 477
QY 481 QLYDHMDIDSMICAGNLQKFG---QDTCCQDSGGPLTCEKDGTVYVYGVISWGLECGKR 536
Db 478 HLFNKTTITNMLCAGDTRTGQNDVHDACQDSGGPLVGMIDKMTLLGIISWGLGCGQK 537
QY 537 --PGVTVQTKPLNWKATIK 555
Db 538 DVEGIYTKVTNVLNWIQDNMK 558

RESULT 6
TPA MOUSE
ID TPA_MOUSE STANDARD; PRT; 559 AA.
AC P11214; Q91VP2;
DT 01-JUL-1989 (Rel. 11, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Tissue-type plasminogen activator precursor (EC 3.4.21.69) (tpa)
DE (t-PA) (t-plasminogen activator).
GN PIAT.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88087303; PubMed=2826484;
RA Rickles R.J., Darrow A.L., Strickland S.;
RT "Molecular cloning of complementary DNA to mouse tissue plasminogen
RT activator mRNA and its expression during F9 teratocarcinoma cell
RT differentiation";
RL J. Biol. Chem. 263:1563-1569 (1988).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Mammary gland;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeng B., Moore T., Max S.I., Wang J., Hsieh F.,
RA Hopkins R.F., Jordan B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Diatchenko L., Soares M.B., Donald M.F., Casavant T.L., Schetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,

RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahney J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield V.S.N., Krywinski M.I., Skalska U., Smalhus D.E.,
RA Scherch A., Schein J.B., Jones S.J.M., Marra M.A.,
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
CC -!- FUNCTION: Converts the abundant, but inactive, zymogen plasminogen
CC to plasmin by hydrolyzing a single Arg-Val bond in plasminogen. By
CC controlling plasmin-mediated proteolysis, it plays an important
CC role in tissue remodeling and degradation, in cell migration and
CC many other physiological events.
CC -!- CATALYTIC ACTIVITY: Specific cleavage of Arg-|-Val bond in
CC plasminogen to form plasmin.
CC -!- SUBUNIT: Heterodimer of chain A and chain B held by a disulfide
CC bond.
CC -!- SUBCELLULAR LOCATION: Secreted; extracellular.
CC -!- PTM: THE SINGLE CHAIN, ALMOST FULLY ACTIVE ENZYME, CAN BE FURTHER
CC PROCESSED INTO A TWO-CHAIN FULLY ACTIVE FORM BY A CLEAVAGE AFTER
CC ARG-308 CATALYZED BY PLASMIN. TISSUE KALLIKREIN OR FACTOR XA.
CC -!- MISCELLANEOUS: Binds to the kringle structure of the fibrin A
CC chain. Binding to fibrin enhances its catalytic activity.
CC -!- SIMILARITY: Belongs to peptidase family S1.
CC -!- SIMILARITY: Contains 1 EGF-like domain.
CC -!- SIMILARITY: Contains 1 fibronectin type I domain.
CC -!- SIMILARITY: Contains 2 kringle domains.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@sb-sib.ch).
CC
CC EMBL; J03520; AAA04070.1; --
CC EMBL; BC011256; AAH11256.1; --
CC DDB; A29941; A29941.
CC HSPF; P00750; IASH.
CC MROPS; S01.232; --
CC MGD; MGI:97610; Plat.
CC InterPro; IPR009003; Cys_Ser_trypsin.
CC InterPro; IPR006209; EGF-like.
CC InterPro; IPR000083; FibFnctnl.
CC InterPro; IPR006210; IEGF.
CC InterPro; IPR000001; Kringle.
CC InterPro; IPR001254; Peptidase_S1.
CC InterPro; IPR001314; Peptidase_S1A.
CC Pfam; PF00008; EGF; 1.
CC Pfam; PF00039; fn1; 1.
CC Pfam; PF00051; kringle; 2.
CC Pfam; PF00089; trypsin; 1.
CC PRINTS; PR00722; CHYMOTRYPSIN.
CC PRINTS; PR00018; KRINGLE.
CC ProDom; PD000395; Kringle; 2.
CC SMART; SM00181; EGF; 1.
CC SMART; SM00058; FN1; 1.
CC SMART; SM00130; KR; 2.
CC SMART; SM00020; Tryp_Spc; 1.
CC PROSITE; PS00022; EGF_1; 1.
CC PROSITE; PS01186; EGF_2; 1.
CC PROSITE; PS50026; EGF_3; 1.
CC PROSITE; PS01253; FIBRONECTIN_1; 1.
CC PROSITE; PS00021; KRINGLE_1; 2.
CC PROSITE; PS50070; KRINGLE_2; 2.
CC PROSITE; PS50240; TRYPsin_DOM; 1.
CC PROSITE; PS00134; TRYPsin_HIS; 1.
CC PROSITE; PS00135; TRYPsin_SER; 1.
CC Plasminogen activation; Hydrolase; Serine protease; Glycoprotein; KW

[illegible]


```

Query Match      23.5%; Score 741; DB 1; Length 615;
Best Local Similarity 11.3%; Pred. No. 6.4e-45;
Matches 205; Conservative 79; Mismatches 210; Indels 164; Gaps 34;

QY      8 LHVLLMALVGVKACGSLMSLESID-PDW-TPOQDYDSYED-----YN 50
Db       1 MRAILL-----GFLLSLSLSLUSIPWEAPKHKAEHIVLVTVTGECHPFPQ 52
QV      51 QENSTSTLTHAENP---DWYIT-----EDQ-----ADPC-QFNCEHGDCGLV 90

```

Db 53 YHRQLYHKTHKGRPGPQPCWATTENFDQDQGWGYCLEPKVKDKHSCSKHSPQCKGGTCVN 112
Qy 91 HGSSTCSCLAPFSKCKQKQNTCKDNPCRGQCL-----TQSPYVRCYCKH 140
Db 113 MFSGHCLCPQHLTGNHCK-----EKCFEPQLRFFHKNWYRTEQAARCOCKG 165
Qy 141 PYTGSCSQWV-PVCRPNPCQNGATC---SRHKRSKFTCAPOQKFKCEIGSD-DCY 195
Db 166 P--DAHCQLASQACRTNPLGGRGLEVEGHR-----LCHCPVGYTGFCDVDTKASY 218
Qy 196 VGDGYSYGRGMRTVQNHACLYWNSHLLLOENYMFMEDAETHGIGENFRNPDADK 255
Db 219 DGRGLSYRGLARTTISGAPCPQWASEATYR---NVTAEQARNWGLGGHAFCRNPNDIRP 275
Qy 256 MCFIKVTNDKXWECYCDVSACSAQOVAVPEBSPT-----EPSTKL 295
Db 276 WCFV-LNRDLRSLWECDLAQCCTPQAAP---PTPVSRLHVLPLMPAOPAPKPKQPTTRT 331
Qy 296 P-----GFDSCG-KTEIAERKIKYGGFKSTAGKHQWASLQ 332
Db 332 PPQSOTPGALPAKREQPFSLTENGFLSCGQRLKSLSMTRVVGGLVALRGAHPYIAYL 391
Qy 333 SSLPLTISPMOQH-FCGGALHPCWLVTAHCTDIK---TRHLKVVLGQDQLKK-----EE 384
Db 392 -----WGSFCAGSLIAPCWLVTAHCLQDRPAPEDLTVVLGQERNHSCPEQOT 441
Qy 385 PHEQSRVEKIPKYSHYNERDEIPHNDIALKL-KPVDGHCALSKYKVTCLPDGSP-P 442
Db 442 LAVRSYRLHEAFSPVSY-----QHDLLALLRLQEDADGSCALLSPYVQVCLPSGAARP 494
Qy 443 SSGE-CHISGWTGTEGKSRQ-----LIDAKVLIANTLCSRLQYLDHMDSDMTCAGNL 497
Db 495 SETTLCQVAGWG--HQFEAEYASFLQAQVFPFLSLRCSADPVHGSILLPGMLCAGFL 552
Qy 498 QKPGQDTCQDGGSLTCEKDGK---YVYVGVSMGLECGKE--EGVYTVTKTKFLNWK 551
Db 553 E-GTDCACQDGGSLVCDQEAERRLTQGIISGSGCGDRKPKGVTVDVAYLAWIR 610

RESULT 8
TPA_BOVIN
ID TPA_BOVIN STANDARD; PRT; 566 AA.
AC Q28198;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Tissue-type plasminogen activator precursor (EC 3.4.21.68) (tPA)
DE (t-PA) (t-plasminogen activator).
GN PLAT.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RN SEQUENCE FROM N.A.
RC TISSUE=Kidney;
RA Ravn P., Berglund L., Petersen T.E.;
RT "Cloning and characterization of the bovine plasminogen activators uPA
Rt and tPA";
RL Int. Dairy J. 5:605-617(1995).
CC -|- FUNCTION: Converts the abundant, but inactive, zymogen plasminogen
CC to plasmin by hydrolyzing a single Arg-Val bond in plasminogen. By
CC controlling plasmin-mediated proteolysis, it plays an important
CC role in tissue remodeling and degradation, in cell migration and
CC many other physiopathological events.
CC -|- CATALYTIC ACTIVITY: Specific cleavage of Arg-|-Val bond in
CC plasminogen to form plasmin.
CC -|- SUBUNIT: Heterodimer of chain A and chain B held by a disulfide
CC bond.
CC -|- SUBCELLULAR LOCATION: Secreted; extracellular.
CC -|- PTM: THE SINGLE CHAIN, ALMOST FULLY ACTIVE ENZYME, CAN BE FURTHER

CC PROCESSED INTO A TWO-CHAIN FULLY ACTIVE FORM BY A CLEAVAGE AFTER
CC ARG-314 CATALYZED BY PLASMIN, TISSUE KALLIKREIN OR FACTOR XA.
CC -|- MISCELLANEOUS: Binds to the kringle structure of the fibrin A
CC chain. Binding to fibrin enhances its catalytic activity.
CC -|- SIMILARITY: Belongs to peptidase family S1.
CC -|- SIMILARITY: Contains 1 EGF-like domain.
CC -|- SIMILARITY: Contains 1 fibronectin type I domain.
CC -|- SIMILARITY: Contains 2 kringle domains.
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X85800; CAA59795.1; -.
CC HSP; P00750; IRTF.
CC MEROPS; S01.232; -.
CC InterPro: IPR009003; Cys_Ser_trypsin.
CC InterPro: IPR006209; EGF-like.
CC InterPro: IPR000083; Fibinctnl.
CC InterPro: IPR006210; IEGF.
CC InterPro: IPR000001; Kringle.
CC InterPro: IPR001254; Peptidase_S1.
CC InterPro: IPR001314; Peptidase_S1A.
CC Pfam: PF00008; EGF_1.
CC Pfam: PF00039; fn1; 1.
CC Pfam: PF00051; kringle; 2.
CC Pfam: PF00089; trypsin; 1.
CC PRINTS: PR00722; CHYMOTRYPSIN.
CC PROSITE: PR00018; KRINGLE.
CC PRODOM: PD000395; Kringle; 2.
CC SMART: SM00181; EGF; 1.
CC SMART: SM00058; FN1; 1.
CC SMART: SM00130; KR; 2.
CC SMART: SM00020; Tryp_SPC; 1.
CC PROSITE: PS00022; EGF_1; 1.
CC PROSITE: PS01186; EGF_2; 1.
CC PROSITE: PS00026; EGF_3; 1.
CC PROSITE: PS01253; FIBRONECTIN_1; 1.
CC PROSITE: PS00021; KRINGLE_1; 1.
CC PROSITE: PS00070; KRINGLE_2; 2.
CC PROSITE: PS02040; TRYPSIN_DOM; 1.
CC PROSITE: PS00134; TRYPSIN_HIS; 1.
CC PROSITE: PS00135; TRYPSIN_SER; 1.
CC Plasminogen activation: Hydrolase; Serine protease; Glycoprotein;
CC Plasma; Kringle; EGF-like domain; Repeat; Signal.
CC SIGNAL
CC PROPEP 1 21
CC CHAIN 22 33 BY SIMILARITY.
CC CHAIN 34 566 TISSUE-TYPE PLASMINOGEN ACTIVATOR.
CC CHAIN 34 314 TISSUE-TYPE PLASMINOGEN ACTIVATOR A
CC CHAIN 315 566 TISSUE-TYPE PLASMINOGEN ACTIVATOR B
CC CHAIN 40 82 FIBRONECTIN TYPE-I.
CC DOMAIN 83 121 EGF-LIKE.
CC DOMAIN 128 209 KRINGLE 1.
CC DOMAIN 219 300 KRINGLE 2.
CC DOMAIN 315 566 SERINE PROTEASE.
CC ACT_SITE 361 361 CHARGE RELAY SYSTEM.
CC ACT_SITE 410 410 CHARGE RELAY SYSTEM.
CC ACT_SITE 517 517 CHARGE RELAY SYSTEM.
CC DISULFID 42 72 BY SIMILARITY.
CC DISULFID 70 79 BY SIMILARITY.
CC DISULFID 87 98 BY SIMILARITY.
CC DISULFID 92 109 BY SIMILARITY.
CC DISULFID 111 120 BY SIMILARITY.
CC DISULFID 128 209 BY SIMILARITY.
CC DISULFID 149 191 BY SIMILARITY.
CC DISULFID 180 204 BY SIMILARITY.
CC DISULFID 219 300 BY SIMILARITY.

FT	DISULFID	240	282	BY SIMILARITY.
FT	DISULFID	271	295	BY SIMILARITY.
FT	DISULFID	303	434	INTERCHAIN (BY SIMILARITY).
FT	DISULFID	345	362	BY SIMILARITY.
FT	DISULFID	354	423	BY SIMILARITY.
FT	DISULFID	448	523	BY SIMILARITY.
FT	DISULFID	480	496	BY SIMILARITY.
FT	DISULFID	513	541	BY SIMILARITY.
FT	CARBOHYD	153	153	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	487	487	N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ	SEQUENCE	566 AA;	63701 MW;	2EB6BEB4E32276C3 CRC64;
Query Match				
Best Local Similarity 34.3%; Pred. No. 1.7e-44;				
Matches 182; Conservative 68; Mismatches 179; Indels 92; Gaps 22;				
QY	77	QCPNPEHGDC	--LVHGGTFTSC	CLAPFSGNKKQ-KVQNTC-KNPFCG-RQCCLITGSP 131
Db	87	CSEPCFNGTCH	QALYSSDFVQC	PEGFMGLCEIDATATCYQGVAYRTWSTAESG 146
QY	132	PYVRC	-----VKHPYTPG	SCSQVVPVCPNCPQNGATCSRHKRSKFTCACPDQ--- 181
Db	147	A--ECANWNS	SGGAMKPYSGR	-----RPNATSLG--LGNHN-----YCRNPDQDSK 188
QY	182	-----PK	-----GKFC	IGS-----DDCVGDGYSYRGKMNRTVNQHACLYWNSHLL 223
Db	189	PWCYVFKAGY	ISERFCSTPAK	VAEEDGDCYTGNGLAYRGTRSHTKSGASCLPWNSVFL 248
QY	224	LQNYNMFMDA	ETHGHEINFC	NPDPDAEKPCWPIKVTNDKVKWYECVYSACSQDVAY 283
Db	249	TSKIYTAWSN	PALGLGKHCR	NPDPGDAQPWCHV-WKDRLTWBYCDVPQCV----- 301
QY	284	PEESPTSP	TKLPFGFSCG	TEAERKIKRIYGGFKSTAGKHQWQASLQSSLPATISMPQ 343
Db	302	-----TC	GLRQY-KPQ	FKIUGLFLADITSHPWQAI--FVKNRSPGE 342
QY	344	GHC	GGALHPCWV	LTAHCTDIK--TRHLKVLVGGQDLKKEEFHQSFVREKIPKISHY 401
Db	343	RFLCGG	LISSCWYLSA	HCQBERYPFHLLKVLGRTYRLVPGEERQTFEVEKYIIHKEF 402
QY	402	NERDEIPHND	IALKLPVDP	GHCALSKYKVTCLPDGS--PPSGECHISGMGVETCK 459
Db	403	D--DDTYND	IALHLKSD	LSLTCAESAVRTICLPDASLQDPWTCELSGVGXKHSS 460
QY	460	G--SRQLD	ARKVLANT	CNSRLQYDHDIMDSMICAGNLQKPG-----QDTCCGDSGGP 512
Db	461	PFFSERLKEA	RVRLYPSR	CTSOHLEFNRTVTNNMLCAGDTRSGGDHTNLHDAQQGDSGGP 520
QY	513	LTCEK	GTYYVYGV	VSNGLEGGK--PGVYTVTKPLNWK 551
Db	521	LVC	MKNHMTLV	GIISWGLGGRKDVGVTKVNYLDWIR 561

RESULT 9
ID -UT2 DESRO
AC P15638; STANDARD; PRT; 477 AA.
DT 01-APR-1990 (Rel. 14, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Salivary plasminogen activator alpha 2 precursor (EC 3.4.21.68) (DSPA
DE alpha-2) (BAT-PA) (T-plasminogen activator).
OS Desmodus rotundus (Vampire bat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Chiroptera; Microchiroptera; Phyllostomidae;
OC Desmodontinae; Desmodus.
OX NCBI_TaxID=9430;
XN [1]
RN
RP SEQUENCE FROM N.A.
RC TISSUE=Salivary gland;
EX MEDLINE=92039036; PubMed=1937019;
RA Kraetschmar J., Haendler B., Langer G., Boisdol W., Bringmann P.,
RA Alagon A., Donner P., Schleuning W.D.;

"The plasminogen activator family from the salivary gland of the vampire bat Desmodus rotundus: cloning and expression.";
Gene 105:229-237(1991).
[2]
SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RP TISSUE=Salivary gland;
RC MEDLINE=90036867; PubMed=2509450;
RA Gardell S.J., Duong L.T., Diehl R.E., York J.D., Hare T.R.,
RA Register R.B., Jacobs J.W., Dixon R.A.F., Friedman P.A.;
RA "Isolation, characterization, and cDNA cloning of a vampire bat salivary plasminogen activator.";
J. Biol. Chem. 264:17947-17952(1989).
[3]
CHARACTERIZATION
RP MEDLINE=93393059; PubMed1309059;
RA Schleuning W.-D., Alagon A., Boisdol W., Bringmann P., Petri T.,
RA Kraetschmar J., Haendler B., Langer G., Baldus B., Witt W.,
RA Donner P.;
RA "Plasminogen activators from the saliva of Desmodus rotundus (common vampire bat): unique fibrin specificity.";
Ann. N.Y. Acad. Sci. 667:395-403(1992).
CC -!- FUNCTION: Probably essential to support the feeding habits of this exclusively haematophagous animal. Probable potent thrombolytic agent.
CC -!- CATALYTIC ACTIVITY: Specific cleavage of Arg-|-Val bond in plasminogen to form plasmin.
CC -!- ENZYME REGULATION: Activity toward plasminogen is stimulated in the presence of fibrin I.
CC -!- SUBUNIT: Monomer.
CC -!- DOMAIN: The fibronectin type-I domain mediates binding to fibrin, and the kringle domain apparently mediates fibrin-induced stimulation of activity.
CC -!- SIMILARITY: Belongs to peptidase family S1.
CC -!- SIMILARITY: Contains 1 EGF-like domain.
CC -!- SIMILARITY: Contains 1 fibronectin type I domain.
CC -!- SIMILARITY: Contains 1 kringle domain.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; M63988; AAA31593.1; -;
CC EMBL; J05082; AAA31596.1; -;
CC PIR; A34369; A34369.
CC PIR; J05098; J05098.
CC HSP; P98119; 1A51.
CC MEROPS; S01.232; -;
CC InterPro; IPR009003; Cys_ser_trypsin.
CC InterPro; IPR006209; EGF-like.
CC InterPro; IPR000083; Fibractnl.
CC InterPro; IPR006210; IEGF.
CC InterPro; IPR000001; Kringle.
CC InterPro; IPR001254; Peptidase_S1.
CC InterPro; IPR001314; Peptidase_S1A.
CC Pfam; PF00008; EGF; 1.
CC Pfam; PF00039; fnl; 1.
CC Pfam; PF00051; kringle; 1.
CC Pfam; PF00089; trypsin; 1.
CC PRINTS; PR00722; CHYMOTRYPSIN.
CC PRINTS; PR00018; KRINGLE.
CC ProDom; PD000395; Kringle; 1.
CC SMART; SM00181; EGF; 1.
CC SMART; SM00058; FNI; 1.
CC SMART; SM00130; KR; 1.
CC SMART; SM00020; Tryp_Spc; 1.
CC PROSITE; PS00022; EGF_1; 1.
CC PROSITE; PS01186; EGF_2; 1.
CC PROSITE; PS00026; EGF_3; 1.
CC PROSITE; PS01253; FIBRONECTIN_1; 1.

DR PROSITE; PS00021; KRINGLE 1; 1.
DR PROSITE; PS00070; KRINGLE_2; 1.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00344; TRYPSIN_HIS; 1.
DR PROSITE; PS00355; TRYPSIN_SER; 1.
KW Plasminogen activator; Hydrolase; Serine protease; Glycoprotein;
KW Kringle; EGF-like domain; Signal; Multigene family.
FT SIGNAL 1 36 POTENTIAL.
FT CHAIN 37 477 SALIVARY PLASMINOGEN ACTIVATOR ALPHA 2.
FT DOMAIN 40 82 FIBRONECTIN TYPE-I.
FT DOMAIN 83 121 EGF-LIKE.
FT DOMAIN 128 209 KRINGLE.
FT DOMAIN 225 477 SERINE PROTEASE.
FT ACT_SITE 272 272 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 321 321 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 428 428 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 42 72 BY SIMILARITY.
FT DISULFID 70 79 BY SIMILARITY.
FT DISULFID 87 98 BY SIMILARITY.
FT DISULFID 92 109 BY SIMILARITY.
FT DISULFID 111 120 BY SIMILARITY.
FT DISULFID 128 209 BY SIMILARITY.
FT DISULFID 149 191 BY SIMILARITY.
FT DISULFID 180 204 BY SIMILARITY.
FT DISULFID 214 345 BY SIMILARITY.
FT DISULFID 257 373 BY SIMILARITY.
FT DISULFID 265 334 BY SIMILARITY.
FT DISULFID 359 434 BY SIMILARITY.
FT DISULFID 391 407 BY SIMILARITY.
FT DISULFID 424 452 BY SIMILARITY.
FT CARBOHYD 185 185 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 398 398 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CONFLICT 403 403 N -> K (IN REF. 2).
FT CONFLICT 417 417 Y -> H (IN REF. 2).
FT CONFLICT 435 435 M -> R (IN REF. 2).
SQ SEQUENCE 477 AA; 53719 MW; 17486555C0E5077C CRC64;

Query Match
Best Local Similarity 35.3%; Pred. No. 25e-44;
Matches 155; Conservative 74; Mismatches 170; Indels 58; Gaps 16;

QY 115 KDNPCGRQCLTQSPYYR-----CVCKHYTGSCSQVVPV--CRPNPCQNG 162
Db 42 CRDE--KTQMIYQQESWLRPEVRSKRVEHCRCR--GLAQCHTPVVKSCSELRCFNG 95
QY 163 ATCSRHKRSKFTACPDQFKGFCIGSD--DCYVGDGYSYRGKWRNTVQHACLYWNHSH 221
Db 96 GTCWQAASFDFVCCQPKGTGQCEVDTHATCYKQGVYTRGTSTSGAQCIWNHSH 155
QY 222 LLQENYNNFMEDATGIGEHNFRCRNPDADEKPCFIVKTVNDKVKWEYCDVSACSADY 281
Db 156 LLTRTYNGRRSDAITLGLGNHNYCRPNPNNSKPCVY- IKASKFLEFCSPVCSKA-- 212
QY 282 AYPERSPTEPKLPGDFGCKTEIAERKIKRYGGFKSTAGKHPQASLQSSLPITISM 341
Db 213 -----TCGLRKYKEPQLHST--GGLFTDITSHPWQAIFAQ--NRRSS 251
QY 342 PQGHFCGALHPWVLTAAHCHTDIK--TRHLKVLGDQDLKKEEPHEQSFRVEKIFKYS 399
Db 252 GERFLCGILISCVWLTAHCHQERYPPQHLAVLVGRIVKVPKGEQETFEVEKIVHE 311
QY 400 HYNEDBIPNDIALKLKPDVGHCALESKYVTVCLPDGS--FPSSGSECHISGVWVET 457
Db 312 EFD--DPTVNDIALLLKSGSPQCAQESDSVRAICLPEANLQLPDWTCELSGYGKHKS 369
QY 458 GKG--SRQLLDKVKLIANTLNSRQLYDHMIDSMICAGNLQK---PG-QDTCCGDSG 510
Db 370 SSFYSFQLKEGHVRLYPSRCSKTFNKTVTNNMLCAGDTSRGIYPNVHDCQDGS 429
QY 511 GPLTCEKDGYYVYGVISWGLECGKR--PGVYTVTKFLNWKATIK 555
Db 430 GPLVCMNDNHWTLIGIISWVGCGEKPDPGVYTKVTVNLYGWRDNR 476

RESULT 10
URTB DESRO STANDARD; PRT; 431 AA.
ID URTB DESRO
AC P98121; 1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Salivary plasminogen activator beta precursor (EC 3.4.21.68) (DSPA beta).
OS Desmodus rotundus (Vampire bat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Chiroptera; Microchiroptera; Phyllostomidae;
OC Desmodontinae; Desmodus.
OX NCBI_TaxID=9430;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Salivary gland;
RX MEDLINE=92039036; PubMed=1937019;
RA Kraetzschmar J., Haendler B., Langer G., Boidol W., Bringmann P., Alagon A., Donner P., Schleuning W.D.;
RT "The plasminogen activator family from the salivary gland of the vampire bat Desmodus rotundus: cloning and expression.";
RL Gene 105:229-237(1991).
RN [2]
RP CHARACTERIZATION.
RX MEDLINE=9339059; PubMed=1309059;
RA Schleuning W.-D., Alagon A., Boidol W., Bringmann P., Petri T., Kraetzschmar J., Haendler B., Langer G., Baldus B., Witt W., Donner P.;
RT "Plasminogen activators from the saliva of Desmodus rotundus (common vampire bat): unique fibrin specificity.";
RL Ann. N.Y. Acad. Sci. 667:395-403(1992).
CC -!- FUNCTION: Probably essential to support the feeding habits of this exclusively haematophagous animal. Probable potent thrombolytic agent.
CC -!- CATALYTIC ACTIVITY: Specific cleavage of Arg-|-Val bond in plasminogen to form plasmin.
CC -!- SUBUNIT: Monomer.
CC -!- SIMILARITY: Belongs to peptidase family S1.
CC -!- SIMILARITY: Contains 1 EGF-like domain.
CC -!- SIMILARITY: Contains 1 kringle domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M63989; AAA31594.1; -.
DR PIR; J05059; J05059.
DR HSSP; P98119; 1A51.
DR MEROPS; S01_239; -.
DR InterPro; IPR009003; Cys_Ser_trypsin.
DR InterPro; IPR006209; EGF_like.
DR InterPro; IPR006210; IEGF.
DR InterPro; IPR000001; Kringle.
DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR001314; Peptidase_S1A.
DR Pfam; PF00008; EGF; 1.
DR Pfam; PF00051; kringle; 1.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR PRINTS; PR00018; KRINGLE.
DR ProDom; PD000395; Kringle; 1.
DR SMART; SM00181; EGF; 1.
DR SMART; SM00130; KE; 1.
DR SMART; SM00020; Tryp_Spc; 1.
DR PROSITE; PS00022; EGF_1; 1.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS50026; EGF_3; 1.

```
DR PROSITE; PS00021; KRINGLE_1; 1.
DR PROSITE; PS00070; KRINGLE_2; 1.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Plasminogen activation; Hydrolase; Serine protease; Glycoprotein;
KW Kringle; EGF-like domain; Signal; Multigene family.
FT SIGNAL 1 35
FT CHAIN 37 431
FT DOMAIN 37 75
FT DOMAIN 82 163
FT DOMAIN 179 431
FT ACT_SITE 226 226
FT ACT_SITE 275 275
FT ACT_SITE 382 382
FT DISULFID 41 52
FT DISULFID 46 63
FT DISULFID 65 74
FT DISULFID 82 163
FT DISULFID 103 145
FT DISULFID 134 158
FT DISULFID 168 299
FT DISULFID 211 227
FT DISULFID 219 288
FT DISULFID 313 388
FT DISULFID 345 361
FT DISULFID 378 406
FT CARBOHYD 139 139
FT CARBOHYD 352 352
SQ SEQUENCE 431 AA; 48221 MW; 699B5E675B162CBF CRC64;

Query Match
Best Local Similarity 36.0%; Score 720.5; DB 1; Length 431;
Matches 157; Conservative 73; Mismatches 159; Indels 47; Gaps 13;

QY 134 YRCVCKHPYTGPSQVVPVCRPNPCQNGATCSRHKRSKFTACPDQFGKFCETGSD- 192
D 28 YRQLARGSPRAYGCGSEL-----RCFNGGTQWQAASFDFVCCPKGYTGKQCEVDTHA 80
QY 193 DCYVGDGYSYRGKMRNTVNHACLYWNSHLLQNYNMFVEDATHGIGHNCRPDAD 252
D 81 TCYKDGVTYRGTWSTSSGACQCNWNSNLTRETYNGRRSDAITLGLGNHNYCRNPDN 140
QY 253 EKPWFVKTNDKVKWEYCDVSAGSAQDVAYPEESPTEPSTKLPFGFSCGKTEIAERKIK 312
D 141 SKPWCVY-IKASKFILECFSPVCSKA-----TCGLRKYKEQLH 179
QY 313 RIYGFKSTAGKHPQASQSSLPITISMPGHCFGGALHPWCWILTAHCTDIK--TRH 370
D 180 ST-GGLFTDITSHPWQAIFAQ--NRRSSGERFLCGGILSSCWLTAAHCFQERYPQH 236
QY 371 LKVLGDQDLKKEEFHQSFRVEKIFKYSHYNERDEIPHNDIALILKLPVDGHCALSKY 430
D 237 LRVLGRVYRVKPGKEEQTFVEKCIIEEPD--DDTYNDIALLLQKSGSPQCAQESDS 294
QY 431 VRTVCLPDGS--FPSGSECHISGWGVTGTGK--SRQLLDKAVKLIANTLNSQLYDHM 486
D 295 VRAICLPEANLQLPDWTCELSGYGKHSFPYSEQLKEGVRVLPSSRCKTSKFLPKNT 354
QY 487 IDDSMLCAGNLQK----PG-QPTCGDGGPLTCEKDGTYVYGVTVSWGLECGKR--PGV 539
D 355 VTNNMLCAGDTSGEIYPNVHDAQCQDGGPLVCVNDHNVILLGIISWVGCGEKDIPGV 414
QY 540 YTVQTKFLNWKATIK 555
D 415 YTKVTNVLGWIRDNR 430

RESULT 11
ID -FA12_BOVIN
AC P98140;
DT 01-FEB-1996 (Rel. 33, Created)
```

```
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Coagulation factor XII precursor (EC 3.4.21.38) (Hageman factor)
DE (HAF) (Fragment).
GN F12.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OC NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94247282; PubMed=8186251;
RA Shibuya Y., Samba U., Okabe H., Kambara T., Yamamoto T.;
RT "Primary structure of bovine Hageman factor (blood coagulation factor XII): comparison with human and guinea pig molecules.";
RL Biochim. Biophys. Acta 1206:63-70(1994).
RN [2]
RP SEQUENCE OF 10-21; 350-364 AND 525-550.
RX MEDLINE=77182112; PubMed=861210;
RA Fujikawa K., Walsh A.K., Davie W.E.;
RT "Isolation and characterization of bovine factor XII (Hageman factor)".;
RL Biochemistry 16:2270-2278(1977).
CC -!- FUNCTION: Factor XII is a serum glycoprotein that participates in the initiation of blood coagulation, fibrinolysis, and the generation of bradykinin and angiotensin.
CC -!- CATALYTIC ACTIVITY: Cleaves selectively Arg-|-Ile bonds in factor VII to form factor VIII and factor XI, to form factor Xa.
CC -!- PTM: O- AND N-GLYCOSYLATED (BY SIMILARITY).
CC -!- MISCELLANEOUS: Factor XII, prekallikrein, and HMW kininogen form a complex bound to an anionic surface. Prekallikrein is cleaved by factor XII to form kallikrein, which then cleaves factor XII first to alpha-factor XIIa and then to beta-factor XIIa. Alpha-factor XIIa activates factor XI to factor Xa. Bovine factor XII is cleaved only to alpha-factor XIIa as it lacks the trypsin/kallikrein cleavage site.
CC -!- SIMILARITY: Belongs to peptidase family S1.
CC -!- SIMILARITY: Contains 2 EGF-like domains.
CC -!- SIMILARITY: Contains 1 fibronectin type I domain.
CC -!- SIMILARITY: Contains 1 fibronectin type II domain.
CC -!- SIMILARITY: Contains 1 kringle domain.
CC -!- SIMILARITY: Contains 1 kringle domain.
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
CC EMBL; S70164; AAB30804.2; -
CC PIR; S45281; S45281.
CC HSSP; P00763; 1DPO.
CC MEROPS; S01.211; -.
CC InterPro; IPR009003; Cys Ser trypsin.
CC InterPro; IPR06209; EGF_like
CC InterPro; IPR000083; Fibnctnl.
CC InterPro; IPR000562; FN_Type_II.
CC InterPro; IPR06210; IEGF.
CC InterPro; IPR000001; Kringle.
CC InterPro; IPR01254; Peptidase S1.
CC InterPro; IPR001314; Peptidase_S1A.
CC Pfam; PF00008; EGF; 2.
CC Pfam; PF00039; fn1; 1.
CC Pfam; PF00040; fn2; 1.
CC Pfam; PF00051; kringle; 1.
CC Pfam; PF00089; trypsin; 1.
CC PRINTS; PR00722; CHYMOTRYPSIN.
CC PRINTS; PR00013; ENTPEI1.
CC PRINTS; PR00018; KRINGLE.
CC PRODOM; PD000095; FN_Type_II; 1.
```

DR ProDom; PD000395; Kringle; 1.
 DR SMART; SM00181; EGF; 2.
 DR SMART; SM00058; FN1; 1.
 DR SMART; SM00059; FN2; 1.
 DR SMART; SM00130; KR; 1.
 DR SMART; SM00020; Tryp_Spc; 1.
 DR PROSITE; PS00022; EGF_1; 2.
 DR PROSITE; PS00186; EGF_2; FALSE NEG.
 DR PROSITE; PS00026; EGF_3; 2.
 DR PROSITE; PS01253; FIBRONECTIN_1; 1.
 DR PROSITE; PS00023; FIBRONECTIN_2; 1.
 DR PROSITE; PS00021; KRINGLE_1; 1.
 DR PROSITE; PS00070; KRINGLE_2; 1.
 DR PROSITE; PS00240; TRYPSIN_DOM; 1.
 DR PROSITE; PS00134; TRYPSIN_HIS; 1.
 DR PROSITE; PS00135; TRYPSIN_SER; 1.
 KW Glycoprotein; Blood coagulation; Plasma; Kringle; Serine protease;
 KW Hydrolyase; Fibrinolysis; EGF-like domain; Repeat; Zymogen; Signal.
 FT NON TER 1
 FT SIGNAL 1
 FT CHAIN 10 349
 FT CHAIN 350 593
 FT DOMAIN 37 78
 FT DOMAIN 84 121
 FT DOMAIN 123 163
 FT DOMAIN 164 200
 FT DOMAIN 207 287
 FT DOMAIN 297 333
 FT DOMAIN 350 593
 FT ACT SITE 389 399
 FT ACT SITE 438 438
 FT ACT SITE 541 541
 FT DISULFID 88 100
 FT DISULFID 94 109
 FT DISULFID 111 120
 FT DISULFID 125 153
 FT DISULFID 151 160
 FT DISULFID 168 179
 FT DISULFID 173 188
 FT DISULFID 190 199
 FT DISULFID 207 287
 FT DISULFID 230 263
 FT DISULFID 258 292
 FT DISULFID 336 463
 FT DISULFID 374 390
 FT DISULFID 382 452
 FT DISULFID 413 416
 FT DISULFID 479 547
 FT DISULFID 510 526
 FT DISULFID 537 568
 FT CARBOHYD 99
 FT CARBOHYD 241
 FT CARBOHYD 263
 FT CARBOHYD 410 410
 FT SEQUENCE 593 AA; 65148 MW; 721592BA792BD61F CRC64;

Query Match 22.4%; Score 706.5; DB 1; Length 593;
 Best Local Similarity 33.0%; Pred. No. 1.7e-42;
 Matches 191; Conservative 69; Mismatches 183; Indels 135; Gaps 31;
 QY 56 SSTLTAAENPDWY---TEQADPC-QPNPCHEGGDCLVHGSGFTTCCLAPFSGNKCQKV 111
 DB 64 ATPNFEKQORWAYCLEPKVKVDHCKSHNDCQGGTCVMPDGPGRICADHFTGKHCQK- 122
 QY 112 QNTCKDNPGRGQCLITQSPYYR-----CVCKHP--YTGPCSQVY 151
 DB 123 -----EKCF---EPQFFRFFHNEIWHRLBPAGVVKQCQKGPNAQCKPLASQ-- 166
 QY 152 PVCRPNPCQNGATCSR---HKRSKFTACPDQFQKGFCEIG-SDDCY--VGDGYGRGK 205
 DB 167 -VKTNEFLNGDSCLQAEGR-----LCRCAPFAGRLCDVDLKASCYDDRRGLSYRGM 220
 QY 206 MRYTVNOHACLYWNHLLLOENYNMFMEDAETHGIGHNFCRNPDADEKPCWCFIKVTNDK 265

DB 221 AGTTLGAPCQWASEATY---WNVTAEOVLNWLGDHAFCEPNPDNTEPCFI-WKGRD 276
 QY 266 VKWEYCDVSAC--SAQDVAYPEESP-----TEPSTKLPGFDS--CGKTEIAE----- 308
 DB 277 LSNWYCRLLAPCAAAGHEHFLPSPSALQKPESTTTQTPLPSTLTSKWCSPPLASGPGGC 336
 QY 309 -----RK-----IKRIYGGFKSTAGKHEWQASLOSSLPLTISMPQHGFCGCGALHPCWVLT 360
 DB 337 GQRLRWLSSLNRVGVGLVALPAAPIAALYK-----DQHFCASSLAPCWVLT 387
 QY 361 AHCTDIK--TRHLKVVGLGDDLLKKEEFHEQS-----FRVEKIFK---YSHYNER 404
 DB 388 AHCLQNRPAKELTVVLG-----QDRHNSQCEQCOTLAVRDYRLHEAFSPITYQH---- 437
 QY 405 DEIPHNDLALLK-KPVDGHCALSKYVTVCLPD--GSFPSSGE---CHISGMG-VTEIG 458
 DB 438 -----DLALVRLQESADGCCAPSPFVQVCLPSTPAARPAESEAACEVAGHGFEGG 491
 QY 459 KGSRLQDLDAKVKLIANTLNSRLYDHMDMSICAGNLQKPGQDTCQDSSGGPLTCE-- 516
 DB 492 EYSSFLQEAQVPLIDPQRCSAPDVHGAFTQGMCLAGFLE-GGTDACQDSSGGPLVCEDE 550
 QY 517 -KDGTYVYGVISWGLECGKR--PGVYTVQTKELNWK 551
 DB 551 TPERQLILRGIVSWGSGCGNRLKPGVYTDVANYLAWIR 588
 RESULT 12
 URTL_DESRO STANDARD; PRT; 477 AA.
 AC P98119;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Salivary plasminogen activator alpha 1 precursor (EC 3.4.21.68) (DSPA
 alpha-1).
 OS Desmodus rotundus (Vampire bat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Chiroptera; Microchiroptera; Phyllostomidae;
 OC Desmodontinae; Desmodus.
 OX NCBI_TaxID=9430;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Salivary gland;
 RX MEDLINE=92039036; PubMed=1937019;
 RA Kraetzschmar J., Haendler B., Langer G., Boidol W., Bringmann P.,
 Alagon A., Donner P., Schleuning W.D.;
 RT "The plasminogen activator family from the salivary gland of the
 vampire bat Desmodus rotundus: cloning and expression.";
 RL Gene 105:229-237(1991).
 RN [2]
 RP CHARACTERIZATION.
 RC MEDLINE=93393059; PubMed=1309059;
 RX Schleuning W.-D., Alagon A., Boidol W., Bringmann P., Petri T.,
 Kraetzschmar J., Haendler B., Langer G., Baldus B., Witt W.,
 Donner P.;
 RT "Plasminogen activators from the saliva of Desmodus rotundus (common
 vampire bat): unique fibrin specificity.";
 RL Ann. N.Y. Acad. Sci. 667:395-403(1992).
 RN [3]
 RP X-RAY CRYSTALLOGRAPHY (2.9 ANGSTROMS).
 RC TISSUE=Salivary gland;
 RX MEDLINE=98022741; PubMed=9354616;
 RA Renatus M., Stubbs M.T., Huber R., Bringmann P., Donner P.,
 Schleuning W.D., Bode W.;
 RT "Catalytic domain structure of vampire bat plasminogen activator: a
 molecular paradigm for proteolysis without activation cleavage.";
 RL Biochemistry 36:13483-13493(1997).
 CC -!- FUNCTION: Probably essential to support the feeding habits of this
 CC exclusively haematophagous animal. Potent thrombolytic agent.
 CC -!- CATALYTIC ACTIVITY: Specific cleavage of Arg-|-Val bond in
 CC plasminogen to form plasmin.

	N-LINKED (GLCNAC. .).	
	/FTID=CAR_000027.	
	N-LINKED (GLCNAC. .).	
	/FTID=CAR_000028.	

DISULFID	265	334
FT		
DISULFID	369	434
FT		
DISULFID	391	407
FT		
DISULFID	424	452
FT		
CARBOHYD	153	153
FT		
CARBOHYD	398	398
FT		
TURN	214	215
FT		
STRAND	223	224
FT		
TURN	226	227
FT		
STRAND	230	231
FT		
HELIX	234	236
FT		
TURN	238	239
FT		
STRAND	240	245
FT		
STRAND	254	263
FT		
TURN	264	265
FT		
STRAND	266	269
FT		
HELIX	271	273
FT		
TURN	280	282
FT		
STRAND	284	287
FT		
TURN	297	298
FT		
STRAND	300	309
FT		
TURN	311	312
FT		
TURN	315	317
FT		
TURN	319	320
FT		
STRAND	323	328
FT		
STRAND	338	338
FT		
TURN	339	340
FT		
STRAND	341	341
FT		
STRAND	345	345
FT		
TURN	349	350
FT		
TURN	355	356
FT		
STRAND	358	363
FT		
STRAND	366	366
FT		
STRAND	374	374
FT		
STRAND	379	385
FT		
HELIX	388	390
FT		
TURN	393	398
FT		
TURN	403	404
FT		
STRAND	405	409
FT		
TURN	425	426
FT		
TURN	428	429
FT		
STRAND	431	436
FT		
TURN	437	438
FT		
STRAND	439	448
FT		
TURN	455	456
FT		
STRAND	459	463
FT		
HELIX	464	467
FT		
HELIX	468	474
FT		

SQ	SEQUENCE	477 AA;	53616 MW;	AA06FD1739C10E5E CRC64;
----	----------	---------	-----------	-------------------------

Query Match	22.3%;	Score 704.5;	DB 1;	Length 477;
Best Local Similarity	35.9%;	Pred. No. 1.8e-42;		
Matches 161;	Conservative 64;	Mismatches 160;	Indels 63;	Gaps 15

Qy	120	CGRGCLITQSPPYRCVCKHPYTGPSCSQVPV--CRPNPCNGATCSRHKRSKFTCA	177
Dd	72	CDRGA-----RC-----HTVPVNSCSBPCFNGGTCWQAVYDFVCQ	110
Qy	178	CPDPKGKFCBIGS-DDCYVGDSYGYSRGKMRTVNQHACLXNSHLLLOENYNFMDEAE	236
Dd	111	CPAGYTKRCEVDIPATCYEGQVTYRTGTWSTAESRVECINWNSSLTRTYNGRMPDAF	170
Qy	237	THGIEHNFCRNPDADKPWFICKYTNDKVXWECDVSACSQAQVAYEESEPTSTKL P	296
Dd	171	NLGLGNHNYCRNPNGAPKWCYV-IKAGKFTSEGCSPVCSKA-----	212
Qy	297	GFDSGCKTEIARKEIKRIYGFKSTAGKHPMQASIQSSLPITISMFGHFCCGALIHPCW	356
Dd	213	---TCLRKYEQLHST-GGLFTDITSHPQAAIFAQ--NRRSSGERFLCGILISSCW	266
Qy	357	VLTAAHC--TDIKTRHLKVLGDQLKBEETHEQSFVEKIFKYSHYNERDEIPHNDIAL	414

```
Db 267 VITAARHCFQESYLPOLKVLGRVTRVVRKPEEEOQFKVKYIVHKEFD--DDTYNDIAL 324
QY 415 LKLPVVDGHCALSKYKVTCLPDGS--FPPSGECHISGWGTETGK--SRQLLDKAVK 470
Db 325 LQLKSDSPCAQESVSRAICLPEANLQLPDWTCELSGYGKHKSSSPFYSEQLKEGHRV 384
QY 471 LIANTLCNSRQIYDHMIDDSMTACGNLQK----PG-QDTCCGDSGGPLTCEKDTYVYG 525
Db 385 LYPSSRCAPKFLNFKVTNNMLCAGDTSGEYIPNVHDACQDGGGLVCMNDNMHTLLG 444
QY 526 IVSWGLECGKR--PGVYTVQVTKFLNWK 551
Db 445 IISWVGCGEKDPVGYIKVNTYLGWIR 472

RESULT 13
UROK RAT
ID UROK RAT STANDARD; PRT; 432 AA.
AC P29598;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Urokinase-type plasminogen activator precursor (EC 3.4.21.73) (uPA)
DE (U-plasminogen activator)
GN PLAU.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Eischer 344;
RX MEDLINE=92233409; PubMed=1568219;
RA Henderson B.R., Tansey W.P., Phillips S.M., Ramshaw I.A.,
RA Kefferd R.F.;
RT "transcriptional and posttranscriptional activation of urokinase
RT plasminogen activator gene expression in metastatic tumor cells.";
RL Cancer Res. 52:2489-2496(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney;
RA Rabbani S.A.;
RL Submitted (APR-1992) to the EMBL/GenBank/DBJ databases.
CC -! CATALYTIC ACTIVITY: Specific cleavage of Arg-|-Val bond in
CC plasminogen to form plasmin.
CC -! SUBUNIT: Found in high and low molecular mass forms. Each consists
CC of two chains, A and B. The high molecular mass form contains a
CC long chain A. Cleavage occurs after residue 156 in the low
CC molecular mass form to yield a short A1 chain (By similarity).
CC -! SIMILARITY: Belongs to peptidase family S1.
CC -! SIMILARITY: Contains 1 EGF-like domain.
CC -! SIMILARITY: Contains 1 kringle domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X63434; CAA45028.1; -
CC EMBL; X65651; CAA46601.1; -
CC FIR; S24604; S18932.
CC HSSP; P00749; 1KDU.
CC MEROPS; S01.231; -.
CC InterPro; IPR009003; Cys Ser trypsin.
CC InterPro; IPR006209; EGF-like.
CC InterPro; IPR006210; IEGF.
CC InterPro; IPR000001; Kringle.
CC InterPro; IPR006293; Pept_S1A_uPA.
CC InterPro; IPR001254; Peptidase_S1.
```

```
DR InterPro; IPR001314; Peptidase_S1A.
DR Pfam; PF00051; kringle; 1.
DR Pfam; PF00089; trypsin; 1.
DR PIRSF; PIRSF001144; Ufk_plasm_act; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR PRINTS; PR0018; KRINGLE.
DR ProDom; PD000395; Kringle; 1.
DR SMART; SM00181; EGF; 1.
DR SMART; SM00130; KR; 1.
DR SMART; SM00020; Tryp_Spc; 1.
DR PROSITE; PS00022; EGF_1; 1.
DR PROSITE; PS01186; EGF_2; FALSE_NEG.
DR PROSITE; PS00026; EGF_3; 1.
DR PROSITE; PS00021; KRINGLE_1; 1.
DR PROSITE; PS00070; KRINGLE_2; 1.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; FALSE_NEG.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Plasminogen activation; Hydrolase; Serine protease; Glycoprotein;
KW Kringle; EGF-like domain; Zymogen; Signal.
FT SIGNAL 1 19
FT CHAIN 20 432 UROKINASE-TYPE PLASMINOGEN ACTIVATOR.
FT CHAIN 20 177 CHAIN A (BY SIMILARITY).
FT CHAIN 156 177 SHORT A CHAIN (A1) (BY SIMILARITY).
FT CHAIN 179 432 CHAIN B (BY SIMILARITY).
FT DOMAIN 27 63 EGF-LIKE.
FT DOMAIN 70 151 KRINGLE.
FT DOMAIN 152 178 CONNECTING PEPTIDE.
FT DOMAIN 179 432 SERINE PROTEASE.
FT DISULFID 31 39 BY SIMILARITY.
FT DISULFID 33 51 BY SIMILARITY.
FT DISULFID 53 62 INTERCHAIN (BY SIMILARITY).
FT DISULFID 168 300 BY SIMILARITY.
FT DISULFID 210 226 BY SIMILARITY.
FT DISULFID 218 289 BY SIMILARITY.
FT DISULFID 314 383 BY SIMILARITY.
FT DISULFID 346 362 BY SIMILARITY.
FT DISULFID 373 401 BY SIMILARITY.
FT ACT_SITE 225 225 CHARGE RELAY SYSTEM.
FT ACT_SITE 276 276 CHARGE RELAY SYSTEM.
FT ACT_SITE 377 377 CHARGE RELAY SYSTEM.
FT CONFLICT 16 16 N -> H (IN REF. 2).
FT CONFLICT 24 24 E -> G (IN REF. 2).
FT CONFLICT 332 332 D -> N (IN REF. 2).
SQ SEQUENCE 432 AA; 47957 MW; 4EB1B96C716244C8 CRC64;

Query Match 22.3%; Score 704; DB 1; Length 432;
Best Local Similarity 37.9%; Pred. No. 1.8e-42;
Matches 157; Conservative 64; Mismatches 163; Indels 30; Gaps 14;

QY 159 CONGATCSRHKRSKF-TCACPDQFKGKCEIG-SDDCYVDGYSYRGKMNRTVNQHACL 216
Db 33 CONGGVCVSYKYPSSIRRCSPKKFKGEHCETDTSKTYHNGNQSTRGKANTDTKGRPCL 92

QY 217 YWNSHLLQENYMFMEDEATHGIGHNFCRNPDADEKPCWCFIKVTNDKVKWEYCDVSAC 276
Db 93 AWNSPAVLQTYNAHRSDALSGLGKHNYCRNPDNQRPWCYVQIGLKQF-----VQEC 146

QY 277 SAQVAVPESPTPEPTSKL--PGFDSGCKTEIAERIKRIYGGFKSTAGKHQASLQSS 334
Db 147 MVQDCSLSK----KPSSTVDQQGF-QCGQKALRPR--FKIVGGEFTVVENQPPFAIY-- 197

QY 335 LPLTISMPQGHFGGALIHPCWVLTAAHC--TDIKTRHLKXVLGDDQDLKKEEFESQFRV 392
Db 196 LKNKGSGSPSPFKCGSLISPCWVASATHCFVNPQKKEEYVVYLGQSKRNSYNPCEMKFEV 257

QY 393 EKIFKYSHYNERDEIPHNDIALKLPVDGHCALSKYKVTCLPD--GSFPGSGECHIS 450
Db 258 EQLIHEDFSDETLAFHNDIALKIRTSQCCAPQRTIQTICLPFRGDPAPFGDCEIT 317

QY 451 GWGVETGKG--SRQLLDKAVKLIANTLCNSRQIYDHMIDDSMTACGNLQKPGQDTQCG 507
Db 318 GFG-QESATDYFPKDLKMSVKLIISHEQCKQPHYGYSEINVKMLCAADPEWK-TDSCSG 375
```


KW	Kringle; EGF-like domain; Zymogen; Signal.	1	20	BY SIMILARITY.
FT	SIGNAL	21	442	UROKINASE-TYPE PLASMINOGEN ACTIVATOR.
FT	CHAIN	21	188	CHAIN A (BY SIMILARITY).
FT	CHAIN	190	442	CHAIN B (BY SIMILARITY).
FT	DOMAIN	29	65	EGF-LIKE.
FT	DOMAIN	72	153	KRINGLE.
FT	DOMAIN	154	183	CONNECTING PEPTIDE.
FT	DOMAIN	190	442	SERINE PROTEASE.
FT	CARBOHYD	152	152	N-LINKED (GLCNAC. . .).
FT	DISULFID	33	41	BY SIMILARITY.
FT	DISULFID	35	53	BY SIMILARITY.
FT	DISULFID	55	64	BY SIMILARITY.
FT	DISULFID	179	310	INTERCHAIN (BY SIMILARITY).
FT	DISULFID	220	236	BY SIMILARITY.
FT	DISULFID	228	299	BY SIMILARITY.
FT	DISULFID	324	393	BY SIMILARITY.
FT	DISULFID	356	372	BY SIMILARITY.
FT	DISULFID	383	411	BY SIMILARITY.
FT	ACT SITE	235	235	CHARGE RELAY SYSTEM.
FT	ACT SITE	286	286	CHARGE RELAY SYSTEM.
FT	ACT SITE	387	387	CHARGE RELAY SYSTEM.
FT	CONFLICT	241	241	Q > H (IN REF. 1; CAA25806).
FT	CONFLICT	242	242	Q > H (IN REF. 1; CAA26511).
FT	CONFLICT	288	288	A > GS (IN REF. 1; CAA25806).
SO	SEQUENCE	442 AA;	49116 MW;	EE32PCEFS0132IEE CRC64;

Query Match	21.8%;	Score	686;	DB 1;	Length	442;			
Best Local Similarity	37.0%;	Pred. No.	3.3e-41;						
Matches	155;	Conservative	63;	Mismatches	169;	Indels	32;	Gaps	14;

QY	159	QCGATCSRHKRSKF-TCA	CPDQFGKFCBIG-SDDCYVGDGYSYRGKMR	TVNQHACL	216	
DB	35	CLNGKCVSYKYSNTQRC	SPKKFOGEHC	IEDTSQTCPEGNHGSYRGKANTNTG	GRPCL	94
QY	217	YMSHLLQENYMFEDAE	THGIEGHNFQCNPPDADEKPC	FKVTNDKVKWEC	VDYSAC	276
DB	95	PMSATVLLNTYAH	RPDALQLGLGKHYCRNPDNR	PMWCYQVGLKQLVQE-CMP	PNC	153
QY	277	SAOD--VA	PEBSPTEPSKLPDSCGKTEA	ERIKRIYGFKSTAGHKP	QASLQS	333
DB	154	SGGSHRPADYDGN	PFSTPEKVE--FCGQKALR	PR--FKIVGKSTTIENQ	PWFAAI--	207
QY	334	SLPLTISMPG----	HECGGALHPCWVLTAAHC----	TDIKTRHLKVLVDGDLKKE	EFH	386
DB	208	-----YRRHGG	SVTYVCGSLISPCWVVSAT	HCFFINYQKEDI-VYLG	QTLHUS	261
QY	387	EQSPRV	EKIKYSHYNRDEIP	HNIDIALLLKLPVDGHC	ALSKYKVTVC	444
DB	262	EMKEVE	EKLILHEDYSAD	SLAHNDIALKIRTDKGC	QCAQPSRSIQ	321
QY	445	SECHISGW	VTETGK--GSQ	LIDAKYLLANTLCNS	RQLYDHMDSD	502
DB	322	ASCEIV	FGKEDPSDYL	PEQLKTVVYKLSH	RECQPHYGS	380
QY	503	DTCCGDSGG	PLTCEKDG	TVYVYGVISV	MGLECG--KREG	559
DB	381	DSQCGDSG	GPLVCSTQ	GLRTLTLTGIVS	WGREGCA	439

RESULT 15									
UROK_BOVIN	STANDARD;	PRT;	433 AA.						
AC	Q05589;	Q28209;							
DT	01-FEB-1994	(Rel. 28, Created)							
DT	01-FEB-1994	(Rel. 28, Last sequence update)							
DT	28-FEB-2003	(Rel. 41, Last annotation update)							
DE	Urokinase-type plasminogen activator precursor	(EC 3.4.21.73)	(uPA)						
DE	(U-plasminogen activator).								
GN	PLAU.								
OS	Bos taurus (Bovine).								
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;								
OC	Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;								

```

QY 508 DSGGPLTCERKDGTVYVYGVISWGLECGK--RGVYTVQVTQFLNWKATIKSESG 559
DB 376 DSGGPLICNIDGRPTLSGIVSWGSGCAEKNKPGVYTRVSYFLNWIQSHIGENG 429

```

OC Bovidae; Bovinae; Bos.
ON NCBI_TaxID=9913;
RX MEDLINE=93216119; PubMed=8385052;
RP SEQUENCE FROM N.A.
RA Kraetzschmar J., Haendler B., Kojima S., Rifkin D.B.,
RA Schleuning W.-D.,
RT "Bovine urokinase-type plasminogen activator and its receptor:
RT cloning and induction by retinoic acid.";
RL Gene 125:177-183(1993).
RN (2)
RP SEQUENCE OF 12-433 FROM N.A.
RC TISSUE=Kidney;
RA Ravn P., Berglund L., Petersen T.B.;
RT "Cloning and characterization of the bovine plasminogen activators uPA
RT and tPA.";
RL Int. Dairy J. 5:605-617(1995).
CC -!- CATALYTIC ACTIVITY: Specific cleavage of Arg-|-Val bond in
CC plasminogen to form plasmin.
CC -!- INDUCTION: By retinoic acid.
CC -!- SIMILARITY: Belongs to peptidase family S1.
CC -!- SIMILARITY: Contains 1 EGF-like domain.
CC -!- SIMILARITY: Contains 1 kringle domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L03546; AAA5149.1; -;
CC EMBL; X85801; CAA59796.1; -;
CC PIR; JN0560; JN0560.
CC HSSP; P00749; 1LMW.
CC MEROPS; S01.231; -;
CC InterPro; IPR009003; Cys Ser trypsin.
CC InterPro; IPR006209; EGF-like.
CC InterPro; IPR000001; Kringle.
CC InterPro; IPR008293; Pept_S1A_uPA.
CC InterPro; IPR001254; Peptidase_S1.
CC InterPro; IPR001314; Peptidase_S1A.
CC Pfam; PF00051; kringle; 1.
CC Pfam; PF00089; trypsin; 1.
CC PIRSP; PIRSF001144; Urk_plasm act; 1.
CC PRINTS; PR00722; CHYMOTRYPSIN.
CC PRINTS; PR00018; KRINGLE.
CC ProDom; PD000395; Kringle; 1.
CC SMART; SM00130; KR; 1.
CC SMART; SM00020; Tryp_Spc; 1.
CC PROSITE; PS00022; EGF_1; 1.
CC PROSITE; PS01186; EGF_2; FALSE_NEG.
CC PROSITE; PS00026; EGF_3; 1.
CC PROSITE; PS00021; KRINGLE_1; 1.
CC PROSITE; PS00070; KRINGLE_2; 1.
CC PROSITE; PS00240; TRYPSIN_DOM; 1.
CC PROSITE; PS00134; TRYPSIN_HIS; 1.
CC PROSITE; PS00135; TRYPSIN_SER; 1.
CC Plasminogen activation; Hydrolase; Serine protease; Glycoprotein;
KW Kringle; EGF-like domain; Signal; Zymogen.
FT SIGNAL 1 20
FT CHAIN 21 433 UROKINASE-TYPE PLASMINOGEN ACTIVATOR.
FT CHAIN 21 179 CHAIN A (BY SIMILARITY).
FT CHAIN 181 433 CHAIN B (BY SIMILARITY).
FT DOMAIN 29 65 EGF-LIKE.
FT DOMAIN 72 153 KRINGLE.
FT DOMAIN 154 180 CONNECTING PEPTIDE.
FT DOMAIN 181 433 SERINE PROTEASE.
FT DISULFID 33 41 BY SIMILARITY.
FT DISULFID 35 53 BY SIMILARITY.
FT DISULFID 55 64 BY SIMILARITY.

FT DISULFID 170 301 INTERCHAIN (BY SIMILARITY).
FT DISULFID 211 227 BY SIMILARITY.
FT DISULFID 219 290 BY SIMILARITY.
FT DISULFID 315 384 BY SIMILARITY.
FT DISULFID 347 363 BY SIMILARITY.
FT DISULFID 374 402 BY SIMILARITY.
FT ACT_SITE 226 226 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 277 277 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 378 378 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT CONFLICT 189 189 A -> T (IN REF. 2).
SQ SEQUENCE 433 AA; 48730 MW; 4DE1B8D4DA7027A CRC64;

Query Match 21.7%; Score 683.5; DB 1; Length 433;
Best Local Similarity 26.7%; Pred. No. 4.9e-41;
Matches 152; Conservative 66; Mismatches 161; Indels 35; Gaps 13;

QY 159 CONGATCSRHRKRSKF-TCACPDQFKGKFCIG-SDDCVGDCGYSGKQNRVTHQACL 216
DB 35 CLNGGKCVTKYKFSNIQRCSCKPKFQGEHCIEDTSKTCYQGNHGSYRGKANDLSGRPCL 94

QY 217 YNNSHLLIQENYNMFMEDAETHGIGHNFCRNPDADEKPCWCFIKVTNDKVKWEYCDVSAC 276
DB 95 AWDSPTVLLKMYHAHRSDAIQLGLGKNYCRNPDNORRWCYVQI-GLKQFVQFCMVQDC 153

QY 277 SAQDVAYPEESPTETKLPGPSCKGTAEAKIKRIYGGFKSTAGKHPWQASLQ---- 332
DB 154 SV-----GKSPSPSPEKEB--FQCGOKALRPR--FKIVGGQVTNAENQFWFAAIYRRHR 203

QY 333 -SSLPTISMPQGHFGGALIHQCWLTAHC--TDIKTRHLKVLGDQDLKKEEFHEQS 389
DB 204 GGSIT-----YLCGSLISPCWVVSATHCFIDHPKKNYIVYLGQSLNSDTGEMQ 255

QY 390 FRVEKIFKYSHYNERDEIPHNDIALKLKFPVDGHGCALESKYKVTCLP--DGSFPGSEC 447
DB 256 FEVEKLILHEDYSAESLAHENDIALLKIRTSRQCAQPSRSIQTICLPPEHDAHSTRRC 315

QY 448 HISGWVTETG--KGSRLLDKAVKLIANTLCNSRQLYDHWDSDMTACGNLQKPGOOTC 505
DB 316 EITGFGKENPSDYRSDDELKMTFVSLVSEVCOQPHYGAETDKMLCAADPQWE-TDSC 374

QY 506 QGDSGGLTCEKDGTYVYVYGVSVGLFCGK--RPGVYTVTKFLNMIKATIKSE 557
DB 375 QGDSGGPLVCTIQGRLLTGLVSWGRDCAMKYKPGVYTVRSKFLPWINTHTRGE 428

Search completed: May 24, 2004, 09:47:26
Job time : 13.5 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 24, 2004, 09:43:35 ; Search time 39 Seconds
(without alignments)
4530.518 Million cell updates/sec

Title: US-09-912-559-3
Perfect score: 3154
Sequence: 1 MFARMSDLHVLLMALVGT.....TQVTKFLNWIKATIKSBSGF 560

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SPTREMBL 25:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phase:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_rvirus:*
- 16: sp_bacteriaph:*
- 17: sp_archaea:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3154	100.0	560	4 Q14520	Q14520 homo sapien
2	2363.5	74.9	517	11 Q8KOD2	Q8KOD2 mus musculus
3	831	26.3	540	13 Q800Y7	Q800Y7 meleagris g
4	829	26.3	653	11 Q8VCS4	Q8VCS4 mus musculus
5	766.5	24.3	516	4 Q8VJ99	Q8VJ99 homo sapien
6	765.5	24.3	562	4 Q86YK8	Q86YK8 homo sapien
7	758.5	24.0	562	6 Q8SQ23	Q8SQ23 sus scrofa
8	737	23.4	564	6 Q8MKB1	Q8MKB1 oryctolagus
9	728	23.1	615	4 Q81Z25	Q81Z25 homo sapien
10	725.5	23.0	616	6 Q97507	Q97507 sus scrofa
11	712	22.6	395	4 Q9BZW1	Q9BZW1 homo sapien
12	694	22.0	597	11 Q35727	Q35727 mus musculus
13	689	21.8	609	11 Q80YCS	Q80YCS mus musculus
14	667.5	21.2	433	6 Q8MIL0	Q8MIL0 oryctolagus
15	666.5	21.1	433	6 Q8MHY7	Q8MHY7 oryctolagus
16	641.5	20.3	810	4 Q15146	Q15146 homo sapien

17	627.5	19.9	454	6	O46506	O46506 papio hamad
18	587.5	18.6	812	11	Q8ROW3	Q8ROW3 rattus norv
19	587	18.6	806	6	O18783	O18783 macropus eu
20	586	18.6	429	13	Q8AVB0	Q8AVB0 brachydanio
21	575	18.2	334	6	O46507	O46507 papio hamad
22	506	16.0	385	5	Q25101	Q25101 herdmantia m
23	468.5	14.9	868	5	Q9YIV3	Q9YIV3 polyandroca
24	465	14.7	716	11	P70521	P70521 rattus norv
25	460	14.6	709	13	Q90ZN6	Q90ZN6 brachydanio
26	456.5	14.5	761	11	Q99JIC8	Q99JIC8 rattus norv
27	455	14.4	716	11	Q91XG8	Q91XG8 mus musculu
28	454	14.4	471	13	Q804X6	Q804X6 gallus gall
29	450.5	14.3	422	4	Q8WVC1	Q8WVC1 homo sapien
30	449	14.2	716	13	Q91691	Q91691 xenopus lae
31	446.5	14.2	461	6	Q95ND7	Q95ND7 pan troglod
32	446	14.1	581	5	Q9XZM7	Q9XZM7 strongyloce
33	444	14.1	710	13	Q91402	Q91402 xenopus lae
34	443	14.0	455	13	Q78Y86	Q78Y86 xenopus lae
35	440	14.0	717	13	P70006	P70006 xenopus lae
36	439.5	13.9	728	11	Q8C9G5	Q8C9G5 mus musculu
37	438.5	13.9	433	13	Q804X5	Q804X5 gallus gall
38	438.5	13.9	855	11	Q9JJI7	Q9JJI7 rattus norv
39	435.5	13.8	709	13	O7ZTN9	O7ZTN9 xenopus lae
40	433.5	13.7	425	13	Q804X7	Q804X7 gallus gall
41	433	13.7	503	13	Q8AYE4	Q8AYE4 brachydanio
42	432.5	13.7	446	11	Q8K3U6	Q8K3U6 rattus norv
43	432.5	13.7	490	11	Q7TN04	Q7TN04 mus musculu
44	432.5	13.7	624	11	Q9DAT3	Q9DAT3 mus musculu
45	432	13.7	1322	5	Q9NJ55	Q9NJ55 anopheles g

ALIGNMENTS

RESULT 1
Q14520
ID Q14520 PRELIMINARY; PRT; 560 AA.
AC Q14520; O00663;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE HGF activator like protein (Hyaluronan binding protein 2).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Kitamura N.;
RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=96425001; PubMed=8827452;
RA Choi-Miura N.H., Tobe T., Sumiya J., Nakano Y., Sano Y., Mazda T., Tomita M.;
RT "Purification and characterization of a novel hyaluronan-binding protein (PHBP) from human plasma: it has three EGF, a kringle and a serine protease domain, similar to hepatocyte growth factor activator.";
RT J. Biochem. 119:1157-1165(1996).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Colon, and Kidney;
RA Strausberg R.;
RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
CC -!- SIMILARITY: CONTAINS 1 KRINGLE DOMAIN.
DR EMBL; D49742; BAA08576.1; -;
DR EMBL; S83182; AAB46909.1; -;
DR EMBL; BC031412; AAH31412.1; -;
DR PIR; JC4795; JC4795.
DR HSSP; P00763; IDPO.
DR MEROPS; S01.033; -;

```

DR Genew; HGNC:4798; HABP2.
DR GO; GO:0005615; C:extracellular space; TAS.
DR GO; GO:0005339; F:glycosaminoglycan binding; TAS.
DR GO; GO:0007155; P:cell adhesion; TAS.
DR InterPro; IPR009003; Cys Ser trypsin.
DR InterPro; IPR006209; EGF-like.
DR InterPro; IPR000001; Kringle.
DR InterPro; IPR001254; Peptidase S1.
DR InterPro; IPR001314; Peptidase_S1A.
DR Pfam; PF00008; EGF_3.
DR Pfam; PF00051; kringle; 1.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR PRINTS; PR00018; KRINGLE.
DR PRODOM; PR000395; Kringle; 1.
DR PROSITE; PS00022; EGF_2; 3.
DR PROSITE; PS01186; EGF_2; 2.
DR PROSITE; PS00021; KRINGLE 2; 1.
DR PROSITE; PS00070; TRYPsin; 1.
DR PROSITE; PS00240; TRYPsin; 1.
DR PROSITE; PS00134; TRYPsin; 1.
DR PROSITE; PS00135; TRYPsin; 1.
DR EGF-like domain; Glycoprotein; Hydrolase; Kringle; Protease;
KW Serine protease.
SQ SEQUENCE 560 AA; 62671 MW; 5CL907230784ACD4 CRC64;

Query Match
Best Local Similarity 100.0%; Score 3154; DB 4; Length 560;
Matches 560; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MFARMSDLHVLMLVKGKTAGCGSLMSLLESLEDPDTPDQDYSDYEDYQNEENTSTLT 60
Db 1 MFARMSDLHVLMLVKGKTAGCGSLMSLLESLEDPDTPDQDYSDYEDYQNEENTSTLT 60

Qy 61 HAENPDWYTTEDQADPCPNCEHGGDCLVHGSTFTCSCLAFFSGNKCQKQNTCKDNP 120
Db 61 HAENPDWYTTEDQADPCPNCEHGGDCLVHGSTFTCSCLAFFSGNKCQKQNTCKDNP 120

Qy 121 GRGQCLITQSPYYRCVKHPYTGSCQVVPVCRPNPCQNGATCSRHRKRSKFTACPD 180
Db 121 GRGQCLITQSPYYRCVKHPYTGSCQVVPVCRPNPCQNGATCSRHRKRSKFTACPD 180

Qy 181 QFKGFCEIGSDDCVVGSGYSYRGKMNRTVNOHACLYWNSHLLLOENYMFEDAEATHGI 240
Db 181 QFKGFCEIGSDDCVVGSGYSYRGKMNRTVNOHACLYWNSHLLLOENYMFEDAEATHGI 240

Qy 241 GEHNFERNPDADKPKWCFIKVTNDKVKWEYCDVSACSAQDVAYPEESTPSTKLPGDS 300
Db 241 GEHNFERNPDADKPKWCFIKVTNDKVKWEYCDVSACSAQDVAYPEESTPSTKLPGDS 300

Qy 301 CGKTEIAERKIKRIYGGFKSTAGKHPWQASLQSLPLTISMPQGHFCGALIHPCWVLT 360
Db 301 CGKTEIAERKIKRIYGGFKSTAGKHPWQASLQSLPLTISMPQGHFCGALIHPCWVLT 360

Qy 361 AHCTDIKTRHLKVGLDQDLKKEEFHEQSFVEKIFKYSYNERDEI PHNDIALLKLPV 420
Db 361 AHCTDIKTRHLKVGLDQDLKKEEFHEQSFVEKIFKYSYNERDEI PHNDIALLKLPV 420

Qy 421 DGHCALESKYKVTCLPDGSPGSECHTSGVGTETGKSGRQLLDKVKLIANTLNSR 480
Db 421 DGHCALESKYKVTCLPDGSPGSECHTSGVGTETGKSGRQLLDKVKLIANTLNSR 480

Qy 481 QLYDHMDISMCAGNLQKPGDTCQDSSGGLTCEKDTYVYVIVSGLECGKRPVY 540
Db 481 QLYDHMDISMCAGNLQKPGDTCQDSSGGLTCEKDTYVYVIVSGLECGKRPVY 540

Qy 541 TQVTKFLNWKATIKSES GF 560
Db 541 TQVTKFLNWKATIKSES GF 560

RESULT 2
Q8KOD2

```

```

ID Q8KOD2 PRELIMINARY; PRT; 517 AA.
AC Q8KOD2.
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney;
RA Strausberg R.;
RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
CC -1- SIMILARITY: CONTAINS 1 KRINGLE DOMAIN.
DR EMBL; BC031775; AAH31775.1;
DR GO; GO:0005509; F:calcium ion binding; IEA.
DR GO; GO:0004263; F:chymotrypsin activity; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0004295; F:trypsin activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR009003; Cys Ser trypsin.
DR InterPro; IPR000742; EGF_2.
DR InterPro; IPR001881; EGF-Ca.
DR InterPro; IPR006209; EGF-like.
DR InterPro; IPR006210; IEGF.
DR InterPro; IPR000001; Kringle.
DR InterPro; IPR001254; Peptidase S1.
DR InterPro; IPR001314; Peptidase_S1A.
DR Pfam; PF00008; EGF_2.
DR Pfam; PF00051; kringle; 1.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR PRINTS; PR00018; KRINGLE.
DR PRODOM; PD000395; Kringle; 1.
DR SMART; SM00181; EGF; 3.
DR SMART; SM00130; KR; 1.
DR SMART; SM00020; Tryp_Spc; 1.
DR PROSITE; PS00022; EGF_1; 3.
DR PROSITE; PS01186; EGF_2; 2.
DR PROSITE; PS00021; KRINGLE 1; 1.
DR PROSITE; PS00070; TRYPsin; 1.
DR PROSITE; PS00240; TRYPsin; 1.
DR PROSITE; PS00134; TRYPsin; 1.
DR PROSITE; PS00135; TRYPsin; 1.
KW Hypothetical protein; EGF-like domain; Glycoprotein; Hydrolase;
KW Kringle; Protease; Serine protease.
SQ SEQUENCE 517 AA; 57326 MW; 3855A42035A5EA59 CRC64;

Query Match
Best Local Similarity 74.9%; Score 2363.5; DB 11; Length 517;
Best Local Similarity 75.1%; Pred. No. 2.6e-209;
Matches 413; Conservative 35; Mismatches 63; Indels 39; Gaps 1;

Qy 10 VLLLMALVGTACGSLMSLLESLEDPDTPDQDYSDYEDYQNEENTSSLTTHAENPDWY 69
Db 6 VLLLMALVGTACGSLMSLLESLEDPDTPDQDYSDYEDYQNEENTSSLTTHAENPDWY 69

Qy 70 TEDQADPCPNCEHGGDCLVHGSTFTCSCLAFFSGNKCQKQNTCKDNPCCRGCLITQ 129
Db 32 -----DFQCNPCEHGGDCLIRGDTSCSCAPFSGSRCTAQNCKDNPVCHGDCCLITQ 86

Qy 130 SPYRCVCVKHPYTGSCQVVPVCRPNPCQNGATCSRHRKRSKFTACPDQKGFCEI 189
Db 87 KHPYRCACPKYPTGPDCKVLPAKPNPCQNGVCSRRHRRSRFTACPDQKGFCEI 146

Qy 190 GSDDCVVGDGYSYRGKMNRTVNOHACLYWNSHLLLOENYMFEDAEATHGIGHNFCRNP 249
Db 147 GPDDCVVGDGYSYRGKMNRTVNOHACLYWNSHLLLOENYMFEDAEATHGIAHNFCRNP 206

Qy 250 DADKPKWCFIKVTNDKVKWEYCDVSACSAQDVAYPEESTPSTKLPGDSCKTIAER 309

```

```

Db 207 DGDHKEWCFKVNSEKVKWEYCDVTVCPVPTPNVPSLELPEVWMLPGFESCGKTEVAEH 266
Qy 310 KIKRIYGGFKSTAGHPKQASLQSSLPITISMPQHFQCGALIHPCWVLTAAHCTDIXTR 369
Db 267 AVKRIYGGFKSTAGHPKQVSLQSLPITISMPQHFQCGALIHPCWVLTAAHCTDINTK 326
Qy 370 HLKVVGLQDGLKEFPFHQSFRVEKIFKYSYHNERDEIPHNDIALLLKLPVDGHCALSK 429
Db 327 HLKVVGLQDGLKXTEHSEQTFRVEKILKYSQYNERDEIPHNDIALLLKLPVGGHCALES 386
Qy 430 VYKTVCLPDGSPFSGSECHISGWVTETCKGSRQLDADAKVLIANTLCSRLQYDHMIDD 489
Db 387 VYKTVCLPDSPFSGTECHISGWVTEIGEGSRQLDADAKVLIANTLCSRLQYDHTIDD 446
Qy 490 SMICAGNLQKPDQTCQDSGGLPCEKDGTYVYVGIWSGLECGKRGVYVTOYTKFLNW 549
Db 447 SMICAGNLQKPSDTCQDSGGLPCEKDGTYVYVGIWSGLECGKRGVYVTOYTKFLNW 506
Qy 550 IKATIKSESG 559
Db 507 IKTVHREAG 516

RESULT 3
Q800Y7
ID Q800Y7 PRELIMINARY; PRT; 540 AA.
AC Q800Y7
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hepatocyte growth factor activator (Fragment).
OS Meleagris gallopavo (Common turkey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Meleagris.
OC NCBI_TaxID=9103;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=2212796; PubMed=12128063;
RA Holsberger D.R., Becker A.E., Thurston R.J., Rice C.D.;
RT "Expression of a hepatocyte growth-factor activator protein in turkey
RT (Meleagris gallopavo) deferent duct epithelial cells.";
RL Comp. Biochem. Physiol. 132:769-777(2002).
RN [2]
RP SEQUENCE FROM N.A.
RA Holsberger D.R., Becker A.E., Thurston R.J., Rice C.D.;
RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY216598; AAC46038.1; -.
DR GO; GO:0005576; C:extracellular; IEA.
DR GO; GO:0004263; F:chymotrypsin activity; IEA.
DR GO; GO:0004295; F:trypsin activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR009003; Cys_Ser_trypsin.
DR InterPro; IPR00742; EGF_2.
DR InterPro; IPR006209; EGF_like.
DR InterPro; IPR000083; Fibinctnl.
DR InterPro; IPR000562; FN_Type_II.
DR InterPro; IPR006210; IEGF.
DR InterPro; IPR000001; Kringle.
DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR001314; Peptidase_S1A.
DR Pfam; PF00008; EGF_2.
DR Pfam; PF00051; kringle; 1.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR PRINTS; PR00018; KRINGLE.
DR ProDom; PD000395; Kringle; 1.
DR SMART; SM00181; EGF_2.
DR SMART; SM00058; FN1; 1.
DR SMART; SM00059; FN2; 1.
DR SMART; SM00130; KR; 1.
DR SMART; SM00020; tryp_Spc; 1.
DR PROSITE; PS00022; EGF_1; 2.
DR PROSITE; PS01186; EGF_2; 1.

```

```

DR PROSITE; PS01253; FIBRONECTIN_1; 1.
DR PROSITE; PS00021; KRINGLE_1; 1.
DR PROSITE; PS50070; KRINGLE_2; 1.
DR PROSITE; PS50240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
FT NON_TER 1
FT NON_TER 540
SQ SEQUENCE 540 AA; 61159 MW; 0BB3B4A89C0B577F CRC64;

Query Match 26.3%; Score 831; DB 13; Length 540;
Best local similarity 34.8%; Pred. No. 7.4e-68;
Matches 188; Conservative 86; Mismatches 208; Indels 58; Gaps 22;

Qy 47 EDYNOQENTSSUTHAENPD--WY--TQDQADQCPNCPCHGGDC-LVHG-STFTCS 98
Db 21 EGKREKDXLCSITHYDRDKWGYCTLLSTSHSDYCANNPCQNGTCTFLAHRRTYHCA 80
Qy 99 CLAPFGKCKQKQVONTCKONPC-----GRGCLITQSPPYVRCVKCKHPYTPGSCS 148
Db 81 CPBEFTGRDQ--MKKCFDDSLYEFDVDMWRSR-----VKQGSVEQCICVDGQM--ECL 131
Qy 149 QVV-PVCRPNPCGATCSRHKRSKFTACPDQKGFCEI--GSDDCVYDGYSGYRGKM 206
Db 132 RVEHKSCVHPCWNGGECCKMIASGKTVCDCKGPFVGKYCNIVPNHHCYRGNGTEYRGA 191
Qy 207 NRTVNOACLYMNSHLILQENYMFEDATHGIGEHNFPCNPDADEKPCWCFIKVINDKV 266
Db 192 KTIISGHSCIPMNSDLILRELVHDSVEKAYQLGLGPFYCRNPDEDEKPCWYI-MKDNSL 250
Qy 267 KWEYCDVSACSAQVAYP--ESPTPESTKLPGFDSGCKTEIAERKIK-RIYGGFKSTAG 323
Db 251 SWEYCNITSCASRRPPVLEDIDTFVPRRP---CGRHKRKSFRVPRILIGSSSLPG 306
Qy 324 KHPQASLQSSLPITISMPQHFQCGALIHPCWVLTAAHC--TDIKTRHLKVVGLQDGLK 381
Db 307 SHPWTAIIYG-----ESFCAGTLIQTCWVSAAHCFANSPKSIKVVGLGHHFN 357
Qy 382 KEEFHEQSFRVEKIFKYSYHNERDEIPHNDIALLLKLPVDGHCALSKYKTVCLPDGS- 440
Db 358 RTDVTQTFEIKYILYPOYSVPRPTEH-DIALIKKNGQRCVAVKSQFVQICLPESNT 416
Qy 441 -FPSGSECHISGWVTE---TGKGSROLLDAKVKLIANTLCSRLQYDHMIDDSMICAGN 496
Db 417 VPDQFKQISGWHKHNITGY-SDVLOETLPIPEEKCRSPETIYGTISENMFCAGY 475
Qy 497 LQKPGQDTCQDSGGLPCEKDGTYVYVGIWSGLECGK--RPGVYVTOYTKFLNWKATI 554
Db 476 FDSK-SDACQDSGGPLACENNEISLYLVISWGDGCGRVNKPQVYTRVPYVNWINERI 534

RESULT 4
Q8VCS4
ID Q8VCS4 PRELIMINARY; PRT; 653 AA.
AC Q8VCS4
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RA Strausberg R.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
CC -!- SIMILARITY: CONTAINS 1 KRINGLE DOMAIN.
DR EMBL; BC019376; AAH19376.1; -.
DR HSSP; P00761; 1ANI.
DR GO; GO:0005576; C:extracellular; IEA.

```

GO; GO:0004263; F:chymotrypsin activity; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0004295; F:trypsin activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR009003; Cys_Ser_trypsin.
DR InterPro; IPR000742; EGF_2_trypsin.
DR InterPro; IPR006209; EGF-like.
DR InterPro; IPR000083; Fibronectin.
DR InterPro; IPR000562; FN_Type_II.
DR InterPro; IPR006210; IEGF.
DR InterPro; IPR000001; Kringle.
DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR001314; Peptidase_S1A.
DR Pfam; PF00008; EGF_2.
DR Pfam; PF00039; fn1.
DR Pfam; PF00040; fn2.
DR Pfam; PF00051; Kringle; 1.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR PRINTS; PR00113; FNTYPEII.
DR PRINTS; PR00018; KRINGLE.
DR PRODOM; PD000995; FN_Type_II; 1.
DR PRODOM; PD000395; Kringle; 1.
DR SMART; SM00181; EGF; 2.
DR SMART; SM00059; FN2; 1.
DR SMART; SM00130; KR; 1.
DR SMART; SM00020; Tryp_Spc; 1.
DR PROSITE; PS00022; EGF_1; 2.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS01233; FIBRONECTIN_1; 1.
DR PROSITE; PS00023; FIBRONECTIN_2; 1.
DR PROSITE; PS00021; KRINGLE_1; 1.
DR PROSITE; PS00070; KRINGLE_2; 1.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hypothetical protein; EGF-like domain; Glycoprotein; Hydrolase;
KW Kringle; Protease; Serine protease.
SQ SEQUENCE 653 AA; 70553 MW; FE18D90174ED6FDD CRC64;

Query Match 26.3%; Score 829; DB 11; Length 653;
Best Local Similarity 34.4%; Pred. No. 1.4e-67;
Matches 190; Conservative 79; Mismatches 187; Indels 96; Gaps 22;

QY 60 THAENPP--WYTEDQ-----DPCQNPCEHGGDLV---HGSTFTCSCLAPFGS 105
DB 133 THNYDRDRAGVCAEVLTPVEGPAILDPCASGCLNGGTCSYHDSGS-YHSCSPLAFTG 191
QY 106 NKCKQVQNTCKD-----NPGCRGQCLITQSPYYRYCYCKRHPYTG 144
DB 192 KDCG--TEKCFDETRYEYFEVGDHWARVSEGHVEQCG---OMEGQA-----RCEDTTH--- 239
QY 145 PSCQVTVCPNPQNGATCSRHKRSKFTACPDQFKGFCEI-GSDCYVGDGYSYR 203
DB 240 -----TACLSPLNGGTHLIVGTGTVCTPLGYAGRFNIVTEHCFLNGTEYR 292
QY 204 GKMRNTVNOHACLWNHLLQENYNNMFEDAEHTGIGENFNRNPDPADKPKCFIKVTN 263
DB 293 GVASTAASGLSCLAWSNLLYQELHVDVSAVAALVGLGPHAYCRNPKDKDERPCYV-VKD 351
QY 264 DKVWEYCDVSAQAQD-----VAYPEESTPEPSTKPLGFDSCGTEAERKIK- 312
DB 352 NALSGWEYCRLTAFCESLARVHSQSPEILALPESAPVRPT-----CGRHKRKRFTFLRP 404
QY 313 RIYGFYKSTAGKHPQASLQSLPLTISMPQGHCGGALHPQWLTAAHC--TDIKTRH 370
DB 405 RIIGSSSLPSHFWLAAIYIG-----NSFCAGSLVHTCWVVSAAHCFANSPPRDS 455
QY 371 LKVLGDQDLKEEFHESQSPVEKIFKYSHYNERDEIPHN-DIALLKLKPVGDHGALESK 429
DB 456 ITVVLGQHFFNRRTDVTQFGIEKYVPYTYLSVFN--PNNHDLVLRLKKGKGERCAVRSQ 513
QY 430 YKTVCLPD--GSFPSSGECHI SGWVTE--TGKGRQLLDAKVKLIANTLCLNSRLDYH 485

DB 514 FVQFICLPAGSSFFPGHKQIAGWGHMDENVSSYNSLLEALVPLVADHKCSSPEVYCA 573
QY 486 MIDDSMICAGNLQKPSQDTCQDSGSLTCBKDGTIVYVGIYSWGLECGK--RPGVYVTOV 543
DB 574 DISPNMLCAGYFDCCK-SDACQDSSGGPLVCEKRGNAVLYGLISWGGCGRLNKPGVYTRV 632
QY 544 TKFLNWKATIK 555
DB 533 ANYVDMINDRIR 644

RESULT 5
Q9BU99
ID Q9BU99 PRELIMINARY; PRT; 516 AA.
AC Q9BU99;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Similar to plasminogen activator, tissue.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
SEQUENCE FROM N.A.
RC TISSUE=Skin;
RA Strauberg R.;
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
CC -1- SIMILARITY: CONTAINS 2 KRINGLE DOMAINS.
DR EMBL; BC002795; AA02795.1;
DR HSP; P00750; 1A5H
DR GO; GO:0004263; F:chymotrypsin activity; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0004295; F:trypsin activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR009003; Cys_Ser_trypsin.
DR InterPro; IPR006209; EGF-like.
DR InterPro; IPR006210; IEGF.
DR InterPro; IPR000001; Kringle.
DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR001314; Peptidase_S1A.
DR Pfam; PF00008; EGF; 1.
DR Pfam; PF00051; Kringle; 2.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR PRINTS; PR00018; KRINGLE.
DR PRODOM; PD000395; Kringle; 2.
DR SMART; SM00181; EGF; 1.
DR SMART; SM00130; KR; 2.
DR SMART; SM00020; Tryp_Spc; 1.
DR PROSITE; PS00022; EGF_1; 1.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS00021; KRINGLE_1; 2.
DR PROSITE; PS00070; KRINGLE_2; 2.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW EGF-like domain; Glycoprotein; Hydrolase; Kringle; Protease;
KW Serine protease.
SQ SEQUENCE 516 AA; 57370 MW; BAB31901FDC96800 CRC64;

Query Match 24.3%; Score 766.5; DB 4; Length 516;
Best Local Similarity 35.0%; Pred. No. 6.2e-62;
Matches 182; Conservative 77; Mismatches 176; Indels 85; Gaps 20;

QY 77 CQNPCEHGGDC--LVHGSTFTCSCLAPFGSGNKCO-KVQNTCKNPF---CGRG----- 123
DB 40 CSEPCFNGGTQQALYFSDFCVQCPEGFAGKCEIDTRATCYEDQGYSYRGWTSTABSG 99
QY 124 -QLITQSPYYRYCYCKRHPYTGSCSQVVPVCRPNPCQ-----NGATCSRHKRSKFTCAC 178

Db 100 AECTNWS-----SALAOKPYSGR-----RPDAIRLGLGNHNYCRNPDNRDPSKPCYV 146
Qy 179 PDQFK-GKF-----CEIGSDDCYVDGYSYRGKMRNVNQHACLWNHLLLOENY 228
Db 147 ---FKAGKYSSEPCSTACSEGNDCYFGNGSAYRGTHSLTESGASCLPWNMILIGKY 203
Qy 229 NMFEDAETHGIGEHNFNCRNPDAEKPCFIKVTNDKVKWEYCDVSACSADQVAYPEEP 288
Db 204 TAQNPQAQALGLGKHNCRNPDGAKPCHV-LKRRLLTWELWEYCDVPSCS-----251
Qy 289 TEPSTKLPGFDSGCKTEIAERKIRIYGKFKSTAGKHPWQASLQSSLPITISMPQGHFCG 348
Db 252 -----TCGLRQYSQPF-RIKGGLFADIASHPWQAIAFAK-HRRSPGERFLCG 297
Qy 349 GALIHPCHWLTAHCTD--IKTRHLKVLGDQDLKKEEFHEOSRVEKIFKYSHYNERDE 406
Db 298 GILISSCWILSAACHCFQBRFPFHLLTVILGTRYVWPGEERQKFEVEKIIVHKEFD--DD 355
Qy 407 IPHNDIALKLKPDVGHCALESKYKVTCLP--DGSFPGSECHISGWTETGKG--SR 462
Db 356 TYNDNIALQLKSDSSRCAQESSVVRTVCLPPADLQLPDWTCELSGVGKHEALSPFYSE 415
Qy 463 QLLDAKVKLIANTLNSRQLYDHMDSDMICAGNLQKPG-----QDTCQDSDGGLPCEK 517
Db 416 RLKEAHVRLYPSRCSQHLNARTVTDNMLCAGDTRSGGPOANLHDACQDSDGGLPVLCLN 475
Qy 518 DGTYYVYGVISWGLECGKR--PGVYTVTKFLNWKATIK 555
Db 476 DGRMTLVGLISWGLCGQKDPGVYTKVTNYLDWIRDNR 515
RESULT 6
Q86YK8
ID Q86YK8 PRELIMINARY; PRT; 562 AA.
AC Q86YK8
DT 01-JUN-2003 (Tremblrel. 24, Created)
DT 01-JUN-2003 (Tremblrel. 24, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Tissue plasminogen activator.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Liu Y., Xu L., Zeng Y., He X.;
RT "cDNA of tissue plasminogen activator."
RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY2211101; RAO34406.1; --
DR GO; GO:0005576; C:extracellular; IEA.
DR GO; GO:0004263; F:chymotrypsin activity; IEA.
DR GO; GO:0004295; F:trypsin activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR009003; Cys Ser trypsin.
DR InterPro; IPR006209; EGF-like.
DR InterPro; IPR000083; Fibronectin.
DR InterPro; IPR000001; Kringle.
DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR001314; Peptidase_S1A.
DR Pfam; PF00008; EGF; 1.
DR Pfam; PF00039; fn1; 1.
DR Pfam; PF00008; EGF; 1.
DR Pfam; PF00051; kringle; 2.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR PRINTS; PR00013; KRINGLE.
DR ProDom; PD000395; Kringle; 2.
DR SMART; SM00058; FN1; 1.
DR SMART; SM00130; KR; 2.
DR SMART; SM00020; Tryp_Spc; 1.
DR PROSITE; PS00022; EGF_1; 1.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS01253; FIBRONECTIN_1; 1.
DR PROSITE; PS00021; KRINGLE_1; 2.

DR PROSITE; PS50070; KRINGLE_2; 2.
DR PROSITE; PS50240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
SQ SEQUENCE 562 AA; 62902 MW; 837D98392F6EDD1F CRC64;
Query Match 24.3%; Score 765.5; DB 4; Length 562;
Best Local Similarity 35.0%; Pred. No. 8.6e-62;
Matches 182; Conservative 76; Mismatches 177; Indels 85; Gaps 20;
Qy 77 CQNPCEHGSDC-LVHGSTFTCSCLAPSPGNKCO-KVQNTCKDNP--CGRG-----123
Db 86 CSEPRCFNGTCCQALYFSDFCQCPGEPAGKCCIDTRATCYEDQGISYRGTWTAESG 145
Qy 124 -OCLITQSPPYKRCVKCHPYTGTPSCSQVVPVCRPNPCQ-----NGATCSRHRKRSKFTCAC 178
Db 146 AECTNWS-----SALAOKPYSGR-----RPDAIRLGLGNHNYCRNPDNRDPSKPCYV 192
Qy 179 PDQFK-GKF-----CEIGSDDCYVDGYSYRGKMRNVNQHACLWNHLLLOENY 228
Db 193 ---FKAGKYSSEPCSTACSEGNDCYFGNGSAYRGTHSLTESGASCLPWNMILIGNY 249
Qy 229 NMFEDAETHGIGEHNFNCRNPDAEKPCFIKVTNDKVKWEYCDVSACSADQVAYPEEP 288
Db 250 TAQNPQAQALGLGKHNCRNPDGAKPCHV-LKRRLLTWELWEYCDVPSCS-----297
Qy 289 TEPSTKLPGFDSGCKTEIAERKIRIYGKFKSTAGKHPWQASLQSSLPITISMPQGHFCG 348
Db 298 -----TCGLRQYSQPF-RIKGGLFADIASHPWQAIAFAK-HRRSPGERFLCG 343
Qy 349 GALIHPCHWLTAHCTD--IKTRHLKVLGDQDLKKEEFHEOSRVEKIFKYSHYNERDE 406
Db 344 GILISSCWILSAACHCFQBRFPFHLLTVILGTRYVWPGEERQKFEVEKIIVHKEFD--DD 401
Qy 407 IPHNDIALKLKPDVGHCALESKYKVTCLP--DGSFPGSECHISGWTETGKG--SR 462
Db 402 TYNDNIALQLKSDSSRCAQESSVVRTVCLPPADLQLPDWTCELSGVGKHEALSPFYSE 461
Qy 463 QLLDAKVKLIANTLNSRQLYDHMDSDMICAGNLQKPG-----QDTCQDSDGGLPCEK 517
Db 462 RLKEAHVRLYPSRCSQHLNARTVTDNMLCAGDTRSGGPOANLHDACQDSDGGLPVLCLN 521
Qy 518 DGTYYVYGVISWGLECGKR--PGVYTVTKFLNWKATIK 555
Db 522 DGRMTLVGLISWGLCGQKDPGVYTKVTNYLDWIRDNR 561
RESULT 7
Q8SQ23
ID Q8SQ23 PRELIMINARY; PRT; 562 AA.
AC Q8SQ23
DT 01-JUN-2002 (Tremblrel. 21, Created)
DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE T-plasminogen activator.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RA TISSUE=Enamel organ;
RA Ding Y., Xue J., Bartlett J.D.;
RT "T-plasminogen activator in tooth tissues."
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: CONTAINS 2 KRINGLE DOMAINS.
DR EMBL; AF364605; AAM00297.1; --
DR HSSP; P00761; IAN1.
DR GO; GO:0005576; C:extracellular; IEA.
DR GO; GO:0004263; F:chymotrypsin activity; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0004295; F:trypsin activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.

```
DR InterPro: IPR009003; Cys_Ser trypsin.
DR InterPro: IPR006209; EGF-like.
DR InterPro: IPR000083; Fibrinectn1.
DR InterPro: IPR006210; IEGF.
DR InterPro: IPR000001; Kringle.
DR InterPro: IPR001254; Peptidase_S1.
DR InterPro: IPR001314; Peptidase_S1A.
DR Pfam: PF00008; EGF_1.
DR Pfam: PF00039; fnl; 1.
DR Pfam: PF00051; kringle; 2.
DR Pfam: PF00089; trypsin; 1.
DR PRINTS: PR00722; CHYMOTRYPSIN.
DR PRINTS: PR00018; KRINGLE.
DR ProDom: PD000395; Kringle; 2.
DR SMART: SM00181; EGF; 1.
DR SMART: SM00058; FNI; 1.
DR SMART: SM00020; Tryp_SPC; 1.
DR SMART: SM00022; EGF_1; 1.
DR PROSITE: PS00022; EGF_2; 1.
DR PROSITE: PS01253; FIBRONECTIN_1; 1.
DR PROSITE: PS00021; KRINGLE_1; 2.
DR PROSITE: PS00070; KRINGLE_2; 2.
DR PROSITE: PS00240; TRYPSIN_DOM; 1.
DR PROSITE: PS00134; TRYPSIN_HIS; 1.
DR PROSITE: PS00135; TRYPSIN_SER; 1.
DR EGF-like domain; Glycoprotein; Hydrolase; Kringle; Protease;
KW Serine protease.
SQ SEQUENCE 562 AA; 63668 MW; F9E5B4C77CB101E8 CRC64;

Query Match 24.0%; Score 758.5; DB 6; Length 562;
Best Local Similarity 33.5%; Pred. No. 3.8e-61;
Matches 183; Conservative 80; Mismatches 201; Indels 83; Gaps 19;

QY 49 YNQEENTSSILTHAENPDWYTDQDPCQPNP-----CEHGDCGL--VHGSGFTCS 98
DB 50 YQHQCWLRLRGRNRVHEHCWMDGQQTQCHSVPVKSCSEPRCFNGTCLQAITFSDFVCQ 109
QY 99 CLAPFSGNKQC-KVQNTCKNP--CGRGCLITQSPFYRCV-----CKHPYTOPSCS 148
DB 110 CPVGFIGQCEIDARATCYEDQGITGTWSTTESGA--ECVNWNTSGLASMEYNGR--- 164
QY 149 QVVPVCPNFCQ-----NGATCSRHKRSKFTC-----ACDPQFKGKEIGSDCYVG 197
DB 165 -----RFDVVKLGLGNHNYCRNPDKDKSPWCYIFKAEKYSDFPCSTPACTKEKECYTG 218
QY 198 DGYSGYRGKARTVQHAACLYWNSHLLQENYMFEDAEHGHGECNFCRNPDADEKPMC 257
DB 219 KGLDYGRTRSLTMSGATCLPWSIVLMGKIYTAWSNAQTLGLGKHNYCRNPDGDTQPMC 278
QY 258 FIKYTNQKVKVEYCDVSACSAQDVAYFEESPTPESTKLPFGDSCGKTEIAERKILYGG 317
DB 279 HV-LKDHKLWYCDLPQCV-----TCGLRQYKEPQF-RIKGG 314
QY 318 FKSTAGKHPWQASLQSLPLTISMPQGHFCGALHPCWLTAAHCTD--IKTRHLKAVVL 375
DB 315 LYADITSHPWQAAL--FVKNRSRGERFLCGGILISSCWLSAAHCFQERFPHVVRVL 372
QY 376 GDQDLKEEFHQFRVEKIFKYSHYNERDIPNDIALKLPVDGHCALSKYKVTVC 435
DB 373 GRTRYLVGEEQAEFEKIVYHKEFD--DDTYNDIALQLKSDSLTCAQESDAVRTVC 430
QY 436 LPDGS--PPSGSECHISGWYTERGKG--SRQLLDKAVKLIANTLCNSRLYDHMIDDSM 491
DB 431 LPEANLQLPDWTCELSGKGHEASSPYSERLKEAHVRLYPSRSTSKHLFNKTIINM 490
QY 492 ICAGNLQKPG-----QDTQCDSGGPLTCEKDGYYVYGVISWGLECGKR--PGVYQVTV 544
DB 491 LCAGTRSGDGNANLHADACQDSGGPLVCMKGNEMTLVGVISWGLGCGQKDPGVYTKVT 550
QY 545 KFLNWK 551
DB 551 NYLNWIR 557

RESULT 8
Q8MKB1 PRELIMINARY; PRT; 564 AA.
ID AC Q8MKB1
AC Q8MKB1; 2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 25, Last annotation update)
DE Tissue-type plasminogen activator.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Euteleostomi; Vertebrata; Euteleostomi;
OC Mammalia; Euteleostomi; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_taxid=9986;
RN [1]
RP SEQUENCE FROM N.A.
RA Sugiki M., Yoshida E., Anai K., Maruyama M.;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
CC -!- SIMILARITY: CONTAINS 2 KRINGLE DOMAINS.
DR EMBL: AY029518; AAK40240.1; .
DR GO: GO:000576; C:extracellular; IEA.
DR GO: GO:004263; F:chymotrypsin activity; IEA.
DR GO: GO:0008233; F:peptidase activity; IEA.
DR GO: GO:0004295; F:trypsin activity; IEA.
DR GO: GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro: IPR009003; Cys_Ser trypsin.
DR InterPro: IPR006209; EGF-like.
DR InterPro: IPR00083; Fibrinectn1.
DR InterPro: IPR006210; IEGF.
DR InterPro: IPR000001; Kringle.
DR InterPro: IPR001254; Peptidase_S1.
DR InterPro: IPR001314; Peptidase_S1A.
DR Pfam: PF00008; EGF; 1.
DR Pfam: PF00039; fnl; 1.
DR Pfam: PF00051; kringle; 2.
DR Pfam: PF00089; trypsin; 1.
DR PRINTS: PR00722; CHYMOTRYPSIN.
DR ProDom: PD000395; Kringle; 2.
DR SMART: SM00181; EGF; 1.
DR SMART: SM00058; FNI; 1.
DR SMART: SM00130; KR; 2.
DR SMART: SM00020; Tryp_SPC; 1.
DR PROSITE: PS00022; EGF_1; 1.
DR PROSITE: PS01253; FIBRONECTIN_1; 1.
DR PROSITE: PS00021; KRINGLE_1; 2.
DR PROSITE: PS00070; KRINGLE_2; 2.
DR PROSITE: PS00240; TRYPSIN_DOM; 1.
DR PROSITE: PS00134; TRYPSIN_HIS; 1.
DR PROSITE: PS00135; TRYPSIN_SER; 1.
DR EGF-like domain; Glycoprotein; Hydrolase; Kringle; Protease;
KW Serine protease.
SQ SEQUENCE 564 AA; 62726 MW; 459DBAC6D4A937C CRC64;

Query Match 23.4%; Score 737; DB 6; Length 564;
Best Local Similarity 35.5%; Pred. No. 3.7e-59;
Matches 182; Conservative 73; Mismatches 180; Indels 78; Gaps 21;

QY 77 CQNPCEHGGDC--LVHGSTTCCLAPFSGNKQC-KVQNTC-KDNPCG-EGQCLITOSP 131
DB 87 CSEPCRLNGTCSQALYFSDFCQPEFGVGRKEVDTRACRYEDRGIGYGTWSTTESG 146
QY 132 PYRVC-----CHPYTGFSCSQVVPVCRPNFCQ-----NGATCSRHKRSKFTCAC-- 178
DB 147 A--QCVWNSSWLAKPYSGR-----KFNALRLGLGNHNYCRNPDRTDKPCYVFR 195
QY 179 PQDFKGFCEI-----GSDCCYVGDGYSGKKNRTVNOHACLYWNSHLLQENYMF 232
DB 196 ACTYSPEFCSTPACSKENKNGCYLKGQAYRGTHTSLTSGASCPLFWNSMLLVGKRYTARQ 255
QY 233 EDAETHGIGEHNFCRNPDADEKFCWCFIKVTNDVKWKEYCDVSACSAQDVAYPESSPTPS 292
```



```

DR PROSITE; PS01186; EGF 2; 1.
DR PROSITE; PS01253; FIBRONECTIN 1; 1.
DR PROSITE; PS00023; FIBRONECTIN 2; 1.
DR PROSITE; PS00021; KRINGLE 1; 1.
DR PROSITE; PS00070; KRINGLE 2; 1.
DR PROSITE; PS00240; TRYPSIN DOM; 1.
DR PROSITE; PS00134; TRYPSIN HIS; 1.
DR PROSITE; PS00135; TRYPSIN SER; 1.
SQ SEQUENCE 615 AA; 67735 MW; 030508870A0C7EDB CRC64;

Query Match      23.1%; Score 728; DB 4; Length 615;
Best Local Similarity 31.1%; Pred. No. 2.8e-58;
Matches 205; Conservative 79; Mismatches 211; Indels 164; Gaps 34;

Qy 8 LHVLLMALVGTACGFSLSMSLESID-PDW-TPDQDYDSYED-----ADPC-QPNPCBHGDCLV 90
Db 1 MRALLLL-----GFLVLSLESTLISIPFWEAPKEHKYKAESHTVVLVTGEPCHFPFQ 52
Qy 51 QEENTSTLTHAENP---DWYTT-----EDQ-----ADPC-QPNPCBHGDCLV 90
Db 53 YHRQLYHKCTHKGRPGQPCWATTNPDQDQRWGYCLEPKVKVDHCKSHSPCKGKGTGVN 112
Qy 91 HGSTFTCSCLAPFSGNKCKQKQNTCKDNPCGRQCL-----ITQSPPYRCVCKH 140
Db 113 MPSPGPHCLCPQHLTGNHCQK-----EKCFEPQLLRFPHKNIWIYRTEQAQAVARCOCKG 165
Qy 141 PYTGSCSOVV-PVCRPNPCNGATC---SRHKRRSKFTCACPDQFKGKFCIGSD-DCY 195
Db 166 P--DAHQORLASQACRTNPNCLHGRCLEVEGHR-----LCHCPVGYTGPFCVDVTRASCY 218
Qy 196 VGDGYSYRGKKNRTVNOHACLYNWSHLLLOENTNMFMEDAEHTGIGEHNFNCRPNDADEKP 255
Db 219 DGRGLSYRGLARTLTSGAPQCPWASEATYR---NVTAEQARNWGLGHAFCRNPDIRP 275
Qy 256 WCFKVTNDVKWEYCDVSACSADQVAYPEEST-----EPSTKL 295
Db 276 WCFV-LNRDRLSWEYCDLQAOQPTTQAAP---PTFVSPLHVLMPAQAPAPPEQPTT 331
Qy 296 P-----GFDSCG-KTIAERKIKRIYGGPKSTAGKHPWASLQ 332
Db 332 PPQSQTPCALPAKREQPPSLTRNGPLSCGQRLKSLSSMTVTVGGLVALGAHPYAAALY 391
Qy 333 SSLPLTISMPQGH-FCCGALIHPCWLVTAACHTDIK--TRHLKVLGLDQDLKK-----EE 384
Db 392 -----WGHSCFAGSJAPCWLVTAACHLQDRPAPEDLTVLQQRNHSCEPCQT 441
Qy 385 FHEQSFVEKIFKYSHYNERDEIPHNDIALKL-KPYDGHCALESKYKVTVCPLDGSF-P 442
Db 442 LAVRSYRLHEAFSPVS-----QHDALLRLQEDADGSCALLSPYVQVCLPSGAARP 494
Qy 443 SGSE-CHISGWVTETGKGRQ-----LLDAVKLIANTLNSRLYDHMDSDMICAGNL 497
Db 495 SETTLQVAGCG--HQFEGAEYASFLQEAQVPLSLERCSAPDVHSGSSILPGMLCAGFL 552
Qy 498 QKPGQDTCQSDSGGLPCEKDGDT---YVYVGIWSWGLECGKR--PGVYTVTKFLNWK 551
Db 553 E-GGTDACQSDSGGLPCEKDGDT---YVYVGIWSWGLECGKR--PGVYTVTKFLNWK 610

RESULT 10
O97507 PRELIMINARY; PRT; 616 AA.
ID O97507;
AC O97507;
DT 01-MAY-1999 (Tremblrel. 10, Created)
DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE FXII.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.

DR PROSITE; PS00022; EGF 1; 2.
DR PROSITE; PS00021; KRINGLE 1; 1.
DR PROSITE; PS00070; KRINGLE 2; 1.
DR PROSITE; PS00240; TRYPSIN DOM; 1.
DR PROSITE; PS00134; TRYPSIN HIS; 1.
DR PROSITE; PS00135; TRYPSIN SER; 1.
SQ SEQUENCE 615 AA; 67735 MW; 030508870A0C7EDB CRC64;

Query Match      23.1%; Score 728; DB 4; Length 615;
Best Local Similarity 31.1%; Pred. No. 2.8e-58;
Matches 205; Conservative 79; Mismatches 211; Indels 164; Gaps 34;

Qy 8 LHVLLMALVGTACGFSLSMSLESID-PDW-TPDQDYDSYED-----ADPC-QPNPCBHGDCLV 90
Db 1 MRALLLL-----GFLVLSLESTLISIPFWEAPKEHKYKAESHTVVLVTGEPCHFPFQ 52
Qy 51 QEENTSTLTHAENP---DWYTT-----EDQ-----ADPC-QPNPCBHGDCLV 90
Db 53 YHRQLYHKCTHKGRPGQPCWATTNPDQDQRWGYCLEPKVKVDHCKSHSPCKGKGTGVN 112
Qy 91 HGSTFTCSCLAPFSGNKCKQKQNTCKDNPCGRQCL-----ITQSPPYRCVCKH 140
Db 113 MPSPGPHCLCPQHLTGNHCQK-----EKCFEPQLLRFPHKNIWIYRTEQAQAVARCOCKG 165
Qy 141 PYTGSCSOVV-PVCRPNPCNGATC---SRHKRRSKFTCACPDQFKGKFCIGSD-DCY 195
Db 166 P--DAHQORLASQACRTNPNCLHGRCLEVEGHR-----LCHCPVGYTGPFCVDVTRASCY 218
Qy 196 VGDGYSYRGKKNRTVNOHACLYNWSHLLLOENTNMFMEDAEHTGIGEHNFNCRPNDADEKP 255
Db 219 DGRGLSYRGLARTLTSGAPQCPWASEATYR---NVTAEQARNWGLGHAFCRNPDIRP 275
Qy 256 WCFKVTNDVKWEYCDVSACSADQVAYPEEST-----EPSTKL 295
Db 276 WCFV-LNRDRLSWEYCDLQAOQPTTQAAP---PTFVSPLHVLMPAQAPAPPEQPTT 331
Qy 296 P-----GFDSCG-KTIAERKIKRIYGGPKSTAGKHPWASLQ 332
Db 332 PPQSQTPCALPAKREQPPSLTRNGPLSCGQRLKSLSSMTVTVGGLVALGAHPYAAALY 391
Qy 333 SSLPLTISMPQGH-FCCGALIHPCWLVTAACHTDIK--TRHLKVLGLDQDLKK-----EE 384
Db 392 -----WGHSCFAGSJAPCWLVTAACHLQDRPAPEDLTVLQQRNHSCEPCQT 441
Qy 385 FHEQSFVEKIFKYSHYNERDEIPHNDIALKL-KPYDGHCALESKYKVTVCPLDGSF-P 442
Db 442 LAVRSYRLHEAFSPVS-----QHDALLRLQEDADGSCALLSPYVQVCLPSGAARP 494
Qy 443 SGSE-CHISGWVTETGKGRQ-----LLDAVKLIANTLNSRLYDHMDSDMICAGNL 497
Db 495 SETTLQVAGCG--HQFEGAEYASFLQEAQVPLSLERCSAPDVHSGSSILPGMLCAGFL 552
Qy 498 QKPGQDTCQSDSGGLPCEKDGDT---YVYVGIWSWGLECGKR--PGVYTVTKFLNWK 551
Db 553 E-GGTDACQSDSGGLPCEKDGDT---YVYVGIWSWGLECGKR--PGVYTVTKFLNWK 610

RESULT 10
O97507 PRELIMINARY; PRT; 616 AA.
ID O97507;
AC O97507;
DT 01-MAY-1999 (Tremblrel. 23, Created)
DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Coagulation factor XII-Mie.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Wada H., Nishio K., Nakatani K., Kasai Y., Abe Y., Nobori T.;
RT "Molecular characterization of coagulation factor XII-Mie.";
RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB095845; BAC23095.1;
DR GO; GO:0005576; C:extracellular; IEA.
DR GO; GO:0005509; F:calcium ion binding; IEA.
DR GO; GO:0004263; F:chymotrypsin activity; IEA.
DR GO; GO:0004295; F:trypsin activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR009003; Cys_Ser_trypsin.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR006209; EGF_like.
DR InterPro; IPR000083; Fibrinctnl.
DR InterPro; IPR000562; FN_Type_II.
DR InterPro; IPR006210; IEGF.
DR InterPro; IPR000001; Kringle.
DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR001314; Peptidase_S1A.
DR Pfam; PF00008; EGF 1.
DR Pfam; PF00039; fn1; 1.
DR Pfam; PF00040; fn2; 1.
DR Pfam; PF00051; kringle; 1.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR PRINTS; PR00013; FNYPEII.
DR PRINTS; PR00018; KRINGLE.
DR ProDom; PD000995; FN_Type_II; 1.
DR ProDom; PD000395; Kringle; 1.
DR SMART; SM00181; EGF; 2.
DR SMART; SM00179; EGF_CA; 2.
DR SMART; SM00058; FN1; 1.
DR SMART; SM00059; FN2; 1.
DR SMART; SM00130; KR; 1.
DR SMART; SM00020; Tryp_Spc; 1.
DR PROSITE; PS00022; EGF_1; 2.
```


Query Match		22.6%; Score 712; DB 4; Length 395;
Best Local Similarity		35.1%; Pred. No. 4.7e-57;
Matches 162; Conservative		67; Mismatches 143; Indels 90; Gaps 16;
QY	114	TKDNPCGRGCLITQSPYRCVKHYTGSCSQVVPVCRPNP-----CQNG-ATCSR 167
DB	3	SCRDE---KTMVYQHQSWLR-----PVLRSNRVEYCMNCGRACQSE 43
QY	168	HKRRSKFTACADQDFKGFCEIGSDCYVGGYSYRGHMTVNOHACLWNHLLQEN 227
DB	44	-----GNSDCYFGNGSAVGRGTHSITSGASCLPWNLSMILICKV 81
QY	228	YNMFMEDAETHGIGHNFNCPNDAKPCFKVTNDKVKWEYCDVSACSADQVAYBEES 287
DB	82	YTAQNPSCAALGLGKHNYCRNPDGAKPWCHV-LKNRLTWEYCDVPSCS----- 130
QY	288	PTEPSTKLPGFDSGCKTETAEKIRIYGGFKSTAGKHPQASLPLTISMPOQH- 346
DB	131	-----TCGLRQYSQGF-RIKGGLFADTASHPWQDAIPAKHE---ESPGERFL 174
QY	347	CGGALIHPCWLVLTAAHCTD--IKRHLKVVLDGDLKKEEFHQSRVEKIFKYSHYNER 404
DB	175	CGGILISSCWILSAAHCFQERPPHLLTVILGRYRVVPGEEQKFECEKIVHKEFD-- 232
QY	405	DEIPHNDIALKLKVPDGHCALESKYKVTCLP--DGSFPGSGECHISGNGVTEGKG-- 460
DB	233	DDTYNDIALQLKSDSSRCAQESSVVRTVCLPPADLQLPDWTCELSGYGKHEALSIFY 292
QY	461	SQLLDAKVLANTLCNSRQLYDHMIDDSMI CAGNLQKPG-----QDTCCGDSGGPLTC 515
DB	293	SERLKEAHVRLPSSRCTSQHLLNRTVTDNMLCAGDTRSGGQANLHACQCGSGPLVC 352
QY	516	EKDGTYYYGVVSWGLECGKR--PGVYTVTKFLNWKATIK 555
DB	353	LNDGRMTLGIISWGLGCGQKDPGVYTVTKVNYLDWIRDNR 394
RESULT 12		
ID	O35727	PRELIMINARY; PRT; 597 AA.
AC	O35727;	
DT	01-JAN-1998	(T=EMBLrel. 05, Created)
DT	01-JAN-1998	(T=EMBLrel. 05, Last sequence update)
DT	01-OCT-2003	(T=EMBLrel. 25, Last annotation update)
DE	Factor XII.	
GN	F12.	
OS	Mus musculus (Mouse).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
OX	NCBI_TaxID=10090;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	TISSUE=Liver;	
RA	Schloesser M., Schwager S., Engel W.;	
RL	Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.	
CC	-1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.	
CC	-1- SIMILARITY: CONTAINS 1 KRINGLE DOMAIN.	
DR	EMBL; X99571; CAA67891.1; --	
DR	HSSP; P00760; 1A07.	
DR	MEKOPS; S01.211; --	
DR	MGD; MGI:1891012; F12.	
DR	GO; GO:0005576; C:extracellular; IEA.	
DR	GO; GO:0004263; F:chymotrypsin activity; IEA.	
DR	GO; GO:0008233; F:peptidase activity; IEA.	
DR	GO; GO:0004295; F:trypsin activity; IEA.	
DR	GO; GO:0006508; P:proteolysis and peptidolysis; IEA.	
DR	InterPro; IPR009003; Cys Ser trypsin.	
DR	InterPro; IPR006209; EGF_like.	
DR	InterPro; IPR000083; Fibnctnl.	
DR	InterPro; IPR000562; FN Type II.	
DR	InterPro; IPR006210; IEGF.	
DR	InterPro; IPR000001; Kringle.	
DR	InterPro; IPR001254; Peptidase_S1.	
InterPro; IPR001314; Peptidase_S1A.		
Pfam; PF00008; EGF; 2.		
Pfam; PF00039; fn1; 1.		
Pfam; PF00040; fn2; 1.		
Pfam; PF00051; kringle; 1.		
Pfam; PF00083; trypsin; 1.		
PRINTS; PR00722; CHYMOTRYP SIN.		
PRINTS; PR00013; FNTYPEII.		
PRINTS; PR00018; KRINGLE.		
ProDom; PD000995; FN Type II; 1.		
ProDom; PD000395; Kringle; 1.		
SMART; SM00181; EGF; 2.		
SMART; SM00058; FN1; 1.		
SMART; SM00059; FN2; 1.		
SMART; SM00130; KR; 1.		
SMART; SM00020; Tryp_SPC; 1.		
PROSITE; PS00022; EGF 1; 2.		
PROSITE; PS01186; EGF 2; 1.		
PROSITE; PS01253; FIBRONECTIN 1; 1.		
PROSITE; PS00023; FIBRONECTIN 2; 1.		
PROSITE; PS00021; KRINGLE 1; 1.		
PROSITE; PS00070; KRINGLE 2; 1.		
PROSITE; PS00240; TRYPSIN DOM; 1.		
PROSITE; PS00134; TRYPSIN HIS; 1.		
PROSITE; PS00135; TRYPSIN SER; 1.		
EGF-like domain; Glycoprotein; Hydrolase; Kringle; Protease;		
Serine protease.		
KW		
SEQUENCE 597 AA; 65638 MW; P3AC07C37D0C0PBA CRC64;		
Query Match		22.0%; Score 694; DB 11; Length 597;
Best Local Similarity		32.8%; Pred. No. 3.7e-55;
Matches 189; Conservative		69; Mismatches 191; Indels 127; Gaps 29;
QY	56	SSILTHAENPDWY--TEDQADPC-QPNPCBHGDCLVHGSTFTCSCLAPFGNKCQKV 111
DB	74	ATTENFDENQMGYCLEPKPKVKDKHCKNPKHGKGTCTNPNGPHCLCPHLTGKCHQK- 132
QY	112	QNTCKNPGCGQLITQSPYY-----RCVCK--HPYTGSCSQVVPVC 154
DB	133	-----EKCFEPQLLKFFHENELWFKTGPVGVARCEKSGSEAHCKPVASQ---AC 178
QY	155	RPNCPQNGATC---SRHKRSKSTCAPDQFKGKCFEIGS-DCVYVDGYSYRGHMTV 210
DB	179	SINPLNGSGCLLAVEDHP-----LCRCPTGYTYGCDLDLWATCYEGRGLSYRGAGTTQ 233
QY	211	NOHACLWNSHLLQENY-NMFMEDAETHGIGHNFNCPNDAKPCFKVTNDKVKWE 269
DB	234	SGAFQCRWT---VEATRYNWKTEQALSWGLGHAFCRNPDNDRPWCFFV-WSGDRLSWD 288
QY	270	YCDVSACSADVA-----YPSPTPE-PTKLPGPS-----CGKTEIAE 308
DB	289	YCGLEQCQQTPTFAPLVVPESQESPSQAPSLSHAPNDSTDHQTSLSTNTMGCGQ---RP 345
QY	309	RK---IKRIYGGFKSTAGKHPQASLQSSLPITISMPOGHFCGGALIHPCWLVTAHCT 364
DB	346	RKGLSSFWRVVGGULVALPGSHFYIALYWG-----NNFCAGSLIAFCWLVTAHCL 396
QY	365	DIK--TRHLKVVLDGDLKKEEFHQ-----SFRVEKIFKYSHYNERDEIHPND 411
DB	397	QNRPAPEELTVLG-----QDRHNQSCWCQTLAVRSYRLHEGFSITY-----QHD 443
QY	412	IALLKL-KPVDGHCALESKYKVTCLDGGSPSGSE--CHISGNGVTEG--KGSROLLD 466
DB	444	LALLRQESKTNSCAILSPHVQVPCLPESGAAPSETVLCVAGHGHLQEGAEYESTLQOE 503
QY	467	AKVKLIANTLCNSRQLYDHMIDDSMI CAGNLQKPGQDTCCGDSGGPLTCCKDGT-----YY 522
DB	504	AQVPFIALLDRCSNHNHGDAILPQMLCAGLE-GGTDACQCGDSGGPLVCB-EGTAERHLT 561
QY	523	VGVISWGLECGKR--PGVYTVTKFLNWKATIKS 556
DB	562	LRGVISWSGCGDRNKPVGIVTDVANYLAWIQKHAS 597


```

DR ProDom; P0000395; Kringle; 1.
DR SMART; SM00130; KR; 1.
DR SMART; SM00020; Tryp_Spc; 1.
DR PROSITE; PS00022; EGF 1; 1.
DR PROSITE; PS00021; KRINGLE 1; 1.
DR PROSITE; PS00070; KRINGLE 2; 1.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
DR PIRSF; PIRSF001144; Urk_plasm_act; 1.
KW Glycoprotein; Hydrolase; Kinase; Kringle; Protease; Serine protease.
SQ SEQUENCE 433 AA; 48375 MW; 65E64F36415549B0 CRC64;

Query Match      21.1%; Score 667.5; DB 6; Length 433;
Best Local Similarity 37.2%; Pred. No. 6.8e-53;
Matches 155; Conservative 62; Mismatches 163; Indels 37; Gaps 14;

QY 159 CONGATCSRHRKRSK-FTCACPDQPKGKFCFEGS-DDCYVGDGYSYRGOMRTVNHACL 216
DB 35 CLNGGTCVTYKFSNIWRCNCPKFKQGEHCIEDTLTKTCYHGDGHSYRGKANTDIMDRPCL 94

QY 217 YNNSHLLLOENYMFEDAEETHGIGHNFCRNPDADKPCFKVTNDKVKWEYCDVSAC 276
DB 95 AANSANVLTKYTHAHRPDALQLGLGHNYCRNPDHQRFPWCIVQGLKQLQGE-CKVHDC 153

QY 277 SAQDVAYPEESPEPTSTKLPFGDSCGKTEIAERKIKRIYGGFKSTAGKHPWQASLQSSLP 336
DB 154 SS-----GKKPALPPGKLE--FQCGQKALRPR--FKIIGGEFTTIENQPF--FAAI----- 198

QY 337 LTISMPQG---HFCGGALHPCWVLTAAHC--TDIKTRHLKVLGDQDLKKEEFHEQSF 390
DB 199 --YRHRGSGVTYVCGSLSPCWVVSATHCFINHKEDYIVYLGSRSLNMTGEMKF 256

QY 391 RVEKIFKYSYNERDEIPNDIALKLKVPDGHCALESKYVKTCLP----DGSPFSGSE 446
DB 257 EVEQLILHEGYRADTLAHHNDIALKILSNNGCQAQPSRSIQITICLPWNADPNF--GTS 314

QY 447 CHISGKGVTTGK--GSRQLLDKVKLIANTLCSNRLYDHMDSDMICAGNLQKPGDGT 504
DB 315 CEITGFGKENSTDYLYPEQLKMTVVKLVSYEQCPHYGSEVTTKMLCAADPQWE-TDS 373

QY 505 CQDSSGGLPCKEKGDTGYVYVYVSWGLECG--KRPQVYTVTKFLNWKATIKSSSG 559
DB 374 CQDSSGGLPVCVQGRMTLTGIVSWGRCALKNKGVTYRVSRLPWRSHRIGEENG 430

RESULT 15
Q8MHY7 PRELIMINARY; PRT; 433 AA.
AC Q8MHY7;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Urokinase-type plasminogen activator.
GN UROKINASE.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RA Yano W.; Yoshida E.; Anai K.; Maruyama M.;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Yano W.; Watanabe M.;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
RN [3]
RP "Oryctolagus cuniculus urokinase-type plasminogen activator, mRNA,
RT complete cds."
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
CC -!- SIMILARITY: CONTAINS 1 KRINGLE DOMAIN.
DR EMBL; AY029517; AAK40239.1; -
DR EMBL; AB087224; BAC02685.1; -

```

```

DR GO; GO:0004263; F:chymotrypsin activity; IEA.
DR GO; GO:0016301; F:kinase activity; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0004295; F:trypsin activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR009003; Cys_ser_trypsin.
DR InterPro; IPR006203; EGF-like.
DR InterPro; IPR000001; Kringle.
DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR001314; Peptidase_S1A.
DR InterPro; IPR008293; Pept_S1A_UPA.
DR Pfam; PF00051; kringle; 1.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR PRINTS; PR00018; KRINGLE.
DR ProDom; PD000395; Kringle; 1.
DR SMART; SM00130; KR; 1.
DR SMART; SM00020; Tryp_Spc; 1.
DR PROSITE; PS00022; EGF 1; 1.
DR PROSITE; PS00021; KRINGLE 1; 1.
DR PROSITE; PS00070; KRINGLE 2; 1.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
DR PIRSF; PIRSF001144; Urk_plasm_act; 1.
KW Glycoprotein; Hydrolase; Kinase; Kringle; Protease; Serine protease.
SQ SEQUENCE 433 AA; 48444 MW; 6DD35A371010A6EE CRC64;

Query Match      21.1%; Score 666.5; DB 6; Length 433;
Best Local Similarity 37.2%; Pred. No. 8.4e-53;
Matches 155; Conservative 61; Mismatches 164; Indels 37; Gaps 14;

QY 159 CONGATCSRHRKRSK-FTCACPDQPKGKFCFEGS-DDCYVGDGYSYRGOMRTVNHACL 216
DB 35 CLNGGTCVTYKFSNIWRCNCPKFKQGEHCIEDTLTKTCYHGDGHSYRGKANTDIMDRPCL 94

QY 217 YNNSHLLLOENYMFEDAEETHGIGHNFCRNPDADKPCFKVTNDKVKWEYCDVSAC 276
DB 95 AANSANVLTKYTHAHRPDALQLGLGHNYCRNPDHQRFPWCIVQGLKQLQGE-CKVHDC 153

QY 277 SAQDVAYPEESPEPTSTKLPFGDSCGKTEIAERKIKRIYGGFKSTAGKHPWQASLQSSLP 336
DB 154 SS-----GKKPALPPGKLE--FQCGQKALRPR--FKIIGGEFTTIENQPF--FAAI----- 198

QY 337 LTISMPQG---HFCGGALHPCWVLTAAHC--TDIKTRHLKVLGDQDLKKEEFHEQSF 390
DB 199 --YRHRGSGVTYVCGSLSPCWVVSATHCFINHKEDYIVYLGSRSLNMTGEMKF 256

QY 391 RVEKIFKYSYNERDEIPNDIALKLKVPDGHCALESKYVKTCLP----DGSPFSGSE 446
DB 257 EVEQLILHEGYRADTLAHHNDIALKILSNNGCQAQPSRSIQITICLPWNADPNF--GTS 314

QY 447 CHISGKGVTTGK--GSRQLLDKVKLIANTLCSNRLYDHMDSDMICAGNLQKPGDGT 504
DB 315 CEITGFGKENSTDYLYPEQLKMTVVKLVSYEQCPHYGSEVTTKMLCAADPQWE-TDS 373

QY 505 CQDSSGGLPCKEKGDTGYVYVYVSWGLECG--KRPQVYTVTKFLNWKATIKSSSG 559
DB 374 CQDSSGGLPVCVQGRMTLTGIVSWGRCALKNKGVTYRVSRLPWRSHRIGEENG 430

```

Search completed: May 24, 2004, 09:48:56
Job time : 41 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 24, 2004, 09:45:05 ; Search time 17.5 Seconds
(without alignments)
1652.031 Million cell updates/sec

Title: US-09-912-559-3
Perfect score: 3154
Sequence: 1 MFARMSDLHVLLMALVYKLT.....TQVTKFLNWKIKSBSGF 560

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/ptodata/2/iaa/5A.COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B.COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A.COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B.COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PTCUS.COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	832.5	26.4	655	1 US-08-148-910-12	Sequence 12, Appl
2	832.5	26.4	655	1 US-08-448-937A-12	Sequence 12, Appl
3	786	24.9	562	6 5244676-5	Patent No. 5244676
4	777.5	24.7	527	6 5520913-1	Patent No. 5520913
5	767.5	24.3	527	2 US-08-811-949-39	Sequence 39, Appl
6	766.5	24.3	527	1 US-07-609-510B-16	Sequence 16, Appl
7	766.5	24.3	527	5 PCT-US91-01025A-2	Sequence 2, Appl
8	766.5	24.3	527	6 5185259-8	Patent No. 5185259
9	766.5	24.3	562	2 US-08-811-949-43	Sequence 43, Appl
10	766.5	24.3	562	2 US-08-560-098A-50	Sequence 50, Appl
11	766.5	24.3	562	2 US-08-883-795A-38	Sequence 38, Appl
12	766.5	24.3	562	4 US-09-703-695A-4	Sequence 4, Appl
13	766.5	24.3	562	6 5185259-3	Patent No. 5185259
14	766.5	24.3	562	6 5200340-2	Patent No. 5200340
15	766.5	24.3	562	6 5344773-2	Patent No. 5344773
16	734.5	23.3	472	2 US-08-811-949-63	Sequence 63, Appl
17	731	23.2	477	2 US-08-560-098A-51	Sequence 51, Appl
18	728	23.1	437	2 US-08-811-949-51	Sequence 51, Appl
19	727	23.1	437	2 US-08-811-949-49	Sequence 49, Appl
20	722	22.9	437	2 US-08-811-949-57	Sequence 57, Appl
21	721	22.9	437	2 US-08-811-949-55	Sequence 55, Appl
22	720	22.8	546	6 5200340-6	Patent No. 5200340
23	715.5	22.7	378	4 US-09-553-498-10	Sequence 10, Appl
24	715.5	22.7	378	4 US-09-618-869-10	Sequence 10, Appl
25	710.5	22.5	355	2 US-08-811-949-47	Sequence 47, Appl
26	709.5	22.5	355	1 US-08-137-116-1	Sequence 1, Appl
27	709.5	22.5	355	1 US-08-217-618-1	Sequence 1, Appl

28	709.5	22.5	355	1 US-08-427-640-2	Sequence 2, Appl
29	709.5	22.5	355	1 US-08-217-617A-1	Sequence 1, Appl
30	709.5	22.5	355	1 US-08-217-616-1	Sequence 1, Appl
31	709.5	22.5	355	2 US-08-811-949-45	Sequence 45, Appl
32	709.5	22.5	355	3 US-08-794-528-1	Sequence 1, Appl
33	709.5	22.5	355	6 5223256-1	Patent No. 5223256
34	709	22.5	347	2 US-08-811-949-1	Sequence 1, Appl
35	707.5	22.4	355	1 US-08-427-640-6	Sequence 6, Appl
36	707.5	22.4	389	2 US-08-811-949-65	Sequence 65, Appl
37	706.5	22.4	355	2 US-08-811-949-59	Sequence 59, Appl
38	703.5	22.3	355	2 US-08-811-949-53	Sequence 53, Appl
39	702.5	22.3	389	2 US-08-811-949-67	Sequence 67, Appl
40	698	22.1	354	2 US-08-811-949-61	Sequence 61, Appl
41	698	22.1	356	1 US-08-427-640-4	Sequence 4, Appl
42	684.5	21.7	356	1 US-08-427-640-8	Sequence 8, Appl
43	663	21.0	430	6 5219569-2	Patent No. 5219569
44	662.5	21.0	411	3 US-09-181-816-1	Sequence 1, Appl
45	661.5	21.0	411	1 US-08-087-163-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-08-148-910-12
; Sequence 12, Application US/08148910
; Patent No. 5466593
; GENERAL INFORMATION:
; APPLICANT: Takeshi SHIMOMURA et al.
; TITLE OF INVENTION: No. 5466593el Protein and Gene Encoding Said Protein
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack
; STREET: 805 Fifteenth Street, N.W., #700
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch,
; MEDIUM TYPE: 500 KB Storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/148,910
; FILING DATE: No. 5466593ember 5, 1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Warren M. Cheek, Jr.
; REGISTRATION NUMBER: 33,367
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-8850
; TELEFAX: 202-371-8856
; TELEX:
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 655 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: human
; US-08-148-910-12

Query Match 26.4%; Score 832.5; DB 1; Length 655;
Best Local Similarity 33.2%; Pred. No. 4e-53;
Matches 192; Conservative 87; Mismatches 216; Indels 83; Gaps 21;

QY 31 ESLDPDWTPOQDYSDYEDYNQEBNTSSTLTHAENPDW-----YYTEDQA----- 74
Db 99 QALTEGDRPCFRFPFRYGRMLHACTSEGSAHRK---WCATTHNYDRDRAGWYCVBEATPPP 155
QY 75 -----DPCQNPCEHGGDC--LVHGSTFTCSCLAPFGSKCKQVQNTCKDNPCGSGOCL 126
Db 156 GGPAALDPCASGCLNGGSCSNTQDPOSYHCSCPRAFTG-----KD--CGTEKCF 203
QY 127 ITQSPPYV-----RCVKCHPYTGSCSQV--PVCRNPCQNGATCSRHK 169
Db 204 DETREYLEGDRWARVRQGHVEQCEC---FGRTWCEGTRHTACLSSPCLNGGTCCHLIV 260
QY 170 RRSKFTACDDQPKGKFCIGSD--CYVGDGYSYRGKQNRVNOHACLWNSHLLLOENY 228
Db 261 ATGTTVCACPPGAFGRCLNTEPDERCFNGTGYRGVASTSASGLSCLWNSDLYQELH 320
QY 229 NMFMEDAETHGIGHNFCRFPDADEKPCWCFIKVTNDKVWEYCDVSACS--QDVAYPBE 286
Db 321 VDSVGAALGLGPHAYCRPNDRPWCYV-VKDSALSWEYCRLEACESLTRVQLSPDL 379
QY 287 SPTPESTKLPGFDCSGKTEIAERKIK-RIYGGFKSTAGKHPQWASLQSSLPITISMPQGH 345
Db 380 LATLPEPASPGRQACGRHRKRTFLRPRIIGSSSLFGSHFWLAAIYIG-----DS 430
QY 346 FCGGALIHPCWLTAAHCTDIKTRH--LKVVLDQDLKKEEFHEQSFVEKIFKYSHYNE 403
Db 431 FCAGSLVHTCWVVSAAHCFSHSPRDSVSVVLGQHFFNRTDVTQTFGIEKYIPYTLISV 490
QY 404 RDEIPHNDIALKLKPVGDHCHALSKYKVTCLPD--GSPFSGSECHISGNG-VTETGKG 460
Db 491 FNPSDH-DLVLRLLKXKGDRCATRSQFVQPCLPPEPGSTFPAGHKCQIAGWGHLDENVSG 549
QY 461 -SRQLLDKVKLIANTLNSRQLYDHMDSDMICAGNLQKPGQDTCOGDSGGPLTCEKDG 519
Db 550 YSSSLREALVPLVADHKCSSPEVYGADISPNMLCAGYFDCK-SDACQDGGGFLACEKNG 608
QY 520 TYVYGVISWGLECGK--RPGVYTVTKFLNWKATIK 555
Db 609 VAYLYGIISWGGCGELHKKPGYVTRVANYVDWINDRIR 646

RESULT 2
US-08-448-937A-12
; Sequence 12, Application US/08448937A
; Patent No. 5677164
; GENERAL INFORMATION:
; APPLICANT: Takeshi SHIMOMURA et al.
; TITLE OF INVENTION: No. 5677164el Protein and Gene Encoding Said Protein
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack
; STREET: 805 Fifteenth Street, N.W., #700
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch,
; MEDIUM TYPE: 500 Kb Storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/448,937A
; FILING DATE: May 24, 1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/148,910
; FILING DATE: No. 5677164ember 5, 1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Warren M. Cheek, Jr.
; REGISTRATION NUMBER: 33,367

REFERENCE/DOCKET NUMBER:
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-371-8850
TELEFAX: 202-371-8856
TELEX:
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 655 amino acids
TYPE: amino acid
STRADEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: human
US-08-448-937A-12

Query Match 26.4%; Score 832.5; DB 1; Length 655;
Best Local Similarity 33.2%; Pred. No. 4e-53;
Matches 192; Conservative 87; Mismatches 216; Indels 83; Gaps 21;

QY 31 ESLDPDWTPOQDYSDYEDYNQEBNTSSTLTHAENPDW-----YYTEDQA----- 74
Db 99 QALTEGDRPCFRFPFRYGRMLHACTSEGSAHRK---WCATTHNYDRDRAGWYCVBEATPPP 155
QY 75 -----DPCQNPCEHGGDC--LVHGSTFTCSCLAPFGSKCKQVQNTCKDNPCGSGOCL 126
Db 156 GGPAALDPCASGCLNGGSCSNTQDPOSYHCSCPRAFTG-----KD--CGTEKCF 203
QY 127 ITQSPPYV-----RCVKCHPYTGSCSQV--PVCRNPCQNGATCSRHK 169
Db 204 DETREYLEGDRWARVRQGHVEQCEC---FGRTWCEGTRHTACLSSPCLNGGTCCHLIV 260
QY 170 RRSKFTACDDQPKGKFCIGSD--CYVGDGYSYRGKQNRVNOHACLWNSHLLLOENY 228
Db 261 ATGTTVCACPPGAFGRCLNTEPDERCFNGTGYRGVASTSASGLSCLWNSDLYQELH 320
QY 229 NMFMEDAETHGIGHNFCRFPDADEKPCWCFIKVTNDKVWEYCDVSACS--QDVAYPBE 286
Db 321 VDSVGAALGLGPHAYCRPNDRPWCYV-VKDSALSWEYCRLEACESLTRVQLSPDL 379
QY 287 SPTPESTKLPGFDCSGKTEIAERKIK-RIYGGFKSTAGKHPQWASLQSSLPITISMPQGH 345
Db 380 LATLPEPASPGRQACGRHRKRTFLRPRIIGSSSLFGSHFWLAAIYIG-----DS 430
QY 346 FCGGALIHPCWLTAAHCTDIKTRH--LKVVLDQDLKKEEFHEQSFVEKIFKYSHYNE 403
Db 431 FCAGSLVHTCWVVSAAHCFSHSPRDSVSVVLGQHFFNRTDVTQTFGIEKYIPYTLISV 490
QY 404 RDEIPHNDIALKLKPVGDHCHALSKYKVTCLPD--GSPFSGSECHISGNG-VTETGKG 460
Db 491 FNPSDH-DLVLRLLKXKGDRCATRSQFVQPCLPPEPGSTFPAGHKCQIAGWGHLDENVSG 549
QY 461 -SRQLLDKVKLIANTLNSRQLYDHMDSDMICAGNLQKPGQDTCOGDSGGPLTCEKDG 519
Db 550 YSSSLREALVPLVADHKCSSPEVYGADISPNMLCAGYFDCK-SDACQDGGGFLACEKNG 608
QY 520 TYVYGVISWGLECGK--RPGVYTVTKFLNWKATIK 555
Db 609 VAYLYGIISWGGCGELHKKPGYVTRVANYVDWINDRIR 646

RESULT 3
5244676-5
; Patent No. 5244676
; APPLICANT: BELL, LESLIE D.;MAYER, ERNEST J.;PALMIER, MARK O.
; TOLUNAY, H.ESER,WARREN, THOMAS G.;WUN, TZB-CHEIN
; TITLE OF INVENTION: MODIFIED TISSUE PLASMINOGEN ACTIVATOR
; WITH MODIFIED GLYCOSYLATION SITE
; NUMBER OF SEQUENCES: 9
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/203,047
; FILING DATE: 06-JUN-1988
; SEQ ID NO:5;

LENGTH: 562
5244676-5

Query Match 24.9%; Score 786; DB 6; Length 562;
Best Local Similarity 33.8%; Pred. No. 8.5e-50;
Matches 181; Conservative 74; Mismatches 177; Indels 104; Gaps 16;
QY 112 QNTCKNPGCGQLITQSPYYR-----CVCKHYTGSCSQVVPV--CRNPPC 159
DB 38 QVICRDE---KTQIYQHQSWLRPLVRSNREVEYWCN---SGRAQCHSVPKSCSPRC 91
QY 160 QNGATCSRHRKRSKFTCAPOFGKCEI--GSDCYVGDGYSYRGKMNRTVNOHACLYW 218
DB 92 FNGGTQQAALYPSDFVCCPGPGAKCEIDGNSDCYFGSGSAYRGTHSTESGASCLPW 151
QY 219 NSHLLQENYMFMEDAETHGIGEHNFRCNPDADKPMCFIKVTNDKVKWEYCDVSACSA 278
DB 152 NSMILIGKVTYTAQNPSAQAALGLGKHNYCRNPDGAKPMCHV--LKNRRLTWECYCDVPCSE 210
QY 279 QD-----VAYPEESTEPSTKL----- 295
DB 211 GNSDCYFNGSAYRGTHSLTESGASCLPWNSMILIGKVTYTAQNPSAQAALGLGKHNYCRN 270
QY 296 -----PFQDSCGKTEIAERKIKIYGGFKSTAGKHPWQASLQ 332
DB 271 DGDAPWCHVLKNRLATWEYCDVPCSTCGLRQYSQPOF--RIKGLPADIASHPWQAIF 329
QY 333 SSLPLTISMPQCHFGGALIHPCWVLTAAHCTD--IKTRHLKVLGDODLKKEEFHEQSF 390
DB 330 AK--HRRSPGFERFLCGGILSSCWILSAHCFQERPPHLLTVILGRYRVVPGEEQKF 387
QY 391 RVEKIKYSHYNERDEIPHNDIALKLPVDGHCALSKYKVTCLP--DGSFPGSGECH 448
DB 388 EVEKIVHKEFD--DDTYNDIALQLKSDSSRCQESSVVRTVCLPAD:QLPDWTECE 445
QY 449 ISGWGTETGKG--SRQLDAKVKLIANTLNSRQLYDHMDSDMI CAGNLQKPG----- 501
DB 446 LSGYKHEALSPFYGERLKEAFVRLYPSRCSCTSOHLNRTVTDNMLCAGDTRSGGPOANL 505
QY 502 QDTCQDGGGGLTCEKDGTYVYVGVSWGLECKGK--PGVTVQTKFLNWKATIK 555
DB 506 HDACQDGGGGLVCLNDGEMTLVGIISWGLGCGQKXDFGVTKVTNYLDWIRDNR 561

RESULT 4

Patent No. 5520913
; APPLICANT: ANDERSON, STEPHEN; BENNETT, WILLIAM F.; BOTSTEIN.
; DAVID; HIGGINS, DEBORAH L.; PAONI, NICHOLAS F.; ZOLLER, MARK J.
; TITLE OF INVENTION: TISSUE PLASMINOGEN ACTIVATOR HAVING
; ZYMOGENIC PROPERTIES
; NUMBER OF SEQUENCES: 35
; APPLICATION DATA:
; FILING DATE: 06-JUL-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 770,510
; FILING DATE: 03-OCT-1991
; APPLICATION NUMBER: 384,608
; FILING DATE: 24-JUL-1989
; APPLICATION NUMBER: 240,856
; FILING DATE: 02-SEP-1988
; SEQ ID NO: 1
; LENGTH: 527
5520913-1

Query Match 24.7%; Score 777.5; DB 6; Length 527;
Best Local Similarity 35.2%; Pred. No. 3.3e-49;
Matches 183; Conservative 76; Mismatches 176; Indels 85; Gaps 20;
QY 77 QCPNCEHGDGC--LVHGTFTTCSCLAPSGNKKQ--KVONTCKNPP--CGRG----- 123
DB 51 CSEPRCFNGGTQQQALYDFDVCQCFEPGAGKCCSIDTRATCYEDQGISYRGTWSTAESG 110

QY 124 -QCLITQSPYYRCVKHPYTGSPSCSQVVPVCRPNPQO-----NGATCSRHRKRSKFTCAC 178
DB 111 ABCTNWN-----SALAKPYSGR-----RPDAIRLGLGNHNYCRNPDSDKPCVY 157
QY 179 PQQFK-GKF-----CEIGSDCYVGDGYSYRGKMNRTVNOHACLYWNSHLLQENY 228
DB 158 ---FRAGKYSSEFCSTPACSEGNSCYFNGSAYRGTHSLTESGASCLPWNSMILRHQKY 214
QY 229 NMFMEDAETHGIGEHNFRCNPDADKPMCFIKVTNDKVKWEYCDVSACSAQDVAYPEESP 288
DB 215 TAQNSAQAALGLGKHNYCRNPDGAKPMCHV--LKNRRLTWECYCDVPCSCS----- 262
QY 289 TEPSTKLPFDSCGKTEIAERKIKIYGGFKSTAGKHPWQASLQSLPLTISMPQCHFGC 348
DB 263 -----TCGLRQYSQPOF--RIKGLPADIASHPWQAIFAK--HRRSPGERHLCG 308
QY 349 GALIHPCWVLTAAHCTD--IKTRHLKVLGDODLKKEEFHEQSFVEKIKYSHYNERDE 406
DB 309 GILISSCWILSAHCFQERPPHLLTVILGRYRVVPGEEQKFVEKIVHKEFD--DD 366
QY 407 IPNDIALKLPVDGHCALSKYKVTCLP--DGSFPGSGECHISGWGTETGKG--SR 462
DB 367 TYDNDIALQLKSDSSRCQESSVVRTVCLPAD:QLPDWTECELSGYKHEALSPFYSE 426
QY 463 QLDDAKVKLIANTLNSRQLYDHMDSDMI CAGNLQKPG-----QDTCQDGGGGLTCEK 517
DB 427 RUKEAHVRLYPSRCSCTSOHLNRTVTDNMLCAGDTRSGGPOANLHDACQDGGGGLVCLN 486
QY 518 DGTYYVGVSWGLECKGK--PGVTVQTKFLNWKATIK 555
DB 487 DGRMTLVGIISWGLGCGQKXDFGVTKVTNYLDWIRDNR 526

RESULT 5

US-08-811-949-39
; Sequence 39, Application US/08811949
; Patent No. 5840533
; GENERAL INFORMATION:
; APPLICANT: NIWA, MINO
; APPLICANT: SAITO, YOSHIMASA
; APPLICANT: SASAKI, HITOSHI
; APPLICANT: HAYASHI, MASAKO
; APPLICANT: NOTANI, JOUJI
; APPLICANT: KOBAYASHI, MASAKAZU
; TITLE OF INVENTION: TISSUE PLASMINOGEN ACTIVATOR
; NUMBER OF SEQUENCES: 67
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
; ADDRESSEE: P.C.
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/811,949
; FILING DATE: 05-MAR-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: OBLON, NORMAN F.
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 18-966-0
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-413-3000
; TELEFAX: 703-413-2220
; INFORMATION FOR SEQ ID NO: 39:
; SEQUENCE CHARACTERISTICS:


```
;
; LENGTH: 527 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-811-949-39

Query Match      24.3%; Score 767.5; DB 2; Length 527;
Best Local Similarity 35.0%; Pred. No. 1.8e-48; Indels 85; Gaps 20;
Matches 182; Conservative 77; Mismatches 176;

QY 77 CQNPCEHGGDC--LVHGSTFTSCCLAPFSGNKCQ-KVQNTCKDNP--CGRG-----123
Db 51 CSEPRCNGGTCCQALYFSDVFCQCEGAGKCEIDTRATCYEDQIGYRGTWSTAEG 110
QY 124 -QCLITQSPYRCVCKHPTGTSCSQVVPVCPNQCQ----NGATCSRHKRSKFTCAC 178
Db 111 AECTNWN-----SALAOKPYSGR-----RPDAIRLGLGNHNYCRNPDROSKPCYV 157
QY 179 PDQFK-GKF-----CEIGSDDCYVGDGYRGKMNRTVNHQACLYWNSHLLQENY 228
Db 158 ---FKAGKYSSEFCSTPACSEGNSDCYFGNGSAYRGTHSLTESGASCLPWNMILIGKY 214
QY 229 NMFMEDAEHTGIGEHNECRNPDADEKPCWCFIKVNDKVWEYCDVSAQDVAYPEESP 288
Db 215 TAQNPSAQAULGLGNHNYCRNPDGDAKPCWCHV-LKNERLTWEYCDVPSCS-----262
QY 289 TEPSTKLPGFDSCKTEIAERKIKRIYGGFKSTAGKHPQASQSSLPITISMPQGHFCG 348
Db 263 -----TCGLRQYSQPF-RIGGLFADIASHPWQAALPAK--HRRSPGERFLCG 308
QY 349 GALHPQWLVTAHCTD--IKTRHLKVLDGQDLKKEEFHEQSPRVEKIFKYSHYNERDE 406
Db 309 GILSSCWILSAAHCFQERPPPHLTVILGRTYRVVPGEEQKFEVEKIVHKEFD--DD 366
QY 407 IPHNDIALKLKPVGDHCALESKYKVTCLP--DGSFPPSGSECHISGWGVTETGK--SR 462
Db 367 TYNDIALQLKSDSSRCQESSVVRTVCLPPADLQLPDWTECELSGYKGHEALSFPYSE 426
QY 463 QLLDAKVKLIANTLCNSRQLYDHMDSDMTCAGNLQKPG-----QDTCQDSSGGPLTCEK 517
Db 427 RLKEARVRLYPSRSCSTQHLLNRTVTDNMLCAGDTRSGGPQANLHDACQDSSGGPLVCLN 486
QY 518 DGTYYVYGVISWGLECKGR--PGVYTOVTKFLNWKATIK 555
Db 487 DGRMTLVGLISWGLGCGCKQKDPGVYTKVTNYLDWIRDNR 526

RESULT 6
PCT-US91-5108-16
; Sequence 16, Application US/076095103
; Patent No. 5326700
; GENERAL INFORMATION:
; APPLICANT: Berg et al.
; TITLE OF INVENTION: Method for Altering Post-Translational Processing of Tissue P
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Eli Lilly and Company
; STREET: Lilly Corporate Center
; CITY: Indianapolis
; STATE: IN.
; COUNTRY: U.S.A.
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.0 Mb storage
; COMPUTER: Macintosh
; OPERATING SYSTEM: Macintosh
; SOFTWARE: Microsoft Word
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/609,510B
; FILING DATE: 19901106
; CLASSIFICATION: 435
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
;
; LENGTH: 527 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-609-510B-16

Query Match      24.3%; Score 766.5; DB 1; Length 527;
Best Local Similarity 35.0%; Pred. No. 2.1e-48; Indels 85; Gaps 20;
Matches 182; Conservative 77; Mismatches 176;

QY 77 CQNPCEHGGDC--LVHGSTFTSCCLAPFSGNKCQ-KVQNTCKDNP--CGRG-----123
Db 51 CSEPRCNGGTCCQALYFSDVFCQCEGAGKCEIDTRATCYEDQIGYRGTWSTAEG 110
QY 124 -QCLITQSPYRCVCKHPTGTSCSQVVPVCPNQCQ----NGATCSRHKRSKFTCAC 178
Db 111 AECTNWN-----SALAOKPYSGR-----RPDAIRLGLGNHNYCRNPDROSKPCYV 157
QY 179 PDQFK-GKF-----CEIGSDDCYVGDGYRGKMNRTVNHQACLYWNSHLLQENY 228
Db 158 ---FKAGKYSSEFCSTPACSEGNSDCYFGNGSAYRGTHSLTESGASCLPWNMILIGKY 214
QY 229 NMFMEDAEHTGIGEHNECRNPDADEKPCWCFIKVNDKVWEYCDVSAQDVAYPEESP 288
Db 215 TAQNPSAQAULGLGNHNYCRNPDGDAKPCWCHV-LKNERLTWEYCDVPSCS-----262
QY 289 TEPSTKLPGFDSCKTEIAERKIKRIYGGFKSTAGKHPQASQSSLPITISMPQGHFCG 348
Db 263 -----TCGLRQYSQPF-RIGGLFADIASHPWQAALPAK--HRRSPGERFLCG 308
QY 349 GALHPQWLVTAHCTD--IKTRHLKVLDGQDLKKEEFHEQSPRVEKIFKYSHYNERDE 406
Db 309 GILSSCWILSAAHCFQERPPPHLTVILGRTYRVVPGEEQKFEVEKIVHKEFD--DD 366
QY 407 IPHNDIALKLKPVGDHCALESKYKVTCLP--DGSFPPSGSECHISGWGVTETGK--SR 462
Db 367 TYNDIALQLKSDSSRCQESSVVRTVCLPPADLQLPDWTECELSGYKGHEALSFPYSE 426
QY 463 QLLDAKVKLIANTLCNSRQLYDHMDSDMTCAGNLQKPG-----QDTCQDSSGGPLTCEK 517
Db 427 RLKEARVRLYPSRSCSTQHLLNRTVTDNMLCAGDTRSGGPQANLHDACQDSSGGPLVCLN 486
QY 518 DGTYYVYGVISWGLECKGR--PGVYTOVTKFLNWKATIK 555
Db 487 DGRMTLVGLISWGLGCGCKQKDPGVYTKVTNYLDWIRDNR 526

RESULT 7
PCT-US91-01025A-2
; Sequence 2, Application PC/TUS9101025A
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; TITLE OF INVENTION: Tissue Plasminogen Activator Having Fibrin
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US91/01025A
; FILING DATE: 19910214
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
```

```
; APPLICATION NUMBER: 07/486,657
; FILING DATE: 1 March 1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Hasak, Janet E.
; REGISTRATION NUMBER: 28,616
; REFERENCE/DOCKET NUMBER: 454P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/266-1896
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 527 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; PCT-US91-01025A-2

Query Match          24.3%; Score 766.5; DB 5; Length 527;
Best Local Similarity 35.0%; Pred. No. 2.1e-48;
Matches 182; Conservative 77; Mismatches 176; Indels 85; Gaps 20;

QY 77 CQNPCEHGGDC--LVHGSTFTCSCLAPFSGNKCO-KVQNTCKDNP--CGRG-----123
Db 51 CSEPRCFNGTCCQALYFSDVFCQCPGFAKCEIDTRATCYEDQGISYRGTWSTAESG 110
QY 124 -QCLITOSPPYRCVCKHPYTGPSCSQVVPVCRPNPCQ-----NGATCSRHKRRSKFTCAC 178
Db 111 AECTNWS-----SALAQKPYSGR-----RPDAIRLGLGNHNYCRNPDNRDSDKPCYV 157
QY 179 PDQFK-GKF-----CEIGSDDCYVGDSYRGKMNRTVNOHACLYNNSHLLQENY 228
Db 158 ---FRAGKYSSEFCSTPACSEGNSDCYFGNGSAYRGTHSLTESGASCLPWNMSMILGKYV 214
QY 229 NMFMEDAETHGIGEHNFERNPDADKPCWFIKVTNDKVKWEYCDVSACSAQDVAYPEESP 288
Db 215 TAQNPQAALGLGKHNYCRNPDGAKPWCHV-LKNRRLTWECYCDVPSCS-----262
QY 289 TEPSTKLPGFDSCKTEIAERKIRYGGFKSTAGKHPWQASLOSPLTITSMPOGHFCG 348
Db 263 -----TCGLRYSQPF-RIKGGLFADIASHPWQAIFAK--HRRSEGERFLCG 308
QY 349 GALIHPCWVLTAACHTD--IKTRHLKVVLGDQDLKKEEFHQSPRVEKIFKYSHYNERDE 406
Db 309 GILLISSCWILSAACHFCQERFPFPHLTVILGRYVVPGEEBQKEVEKYIVHKEFD--DD 366
QY 407 IPHNDIALKLKVPDGHCALESKYVTVCLP--DGSPFSGSECHISGMGVTETGKG--SR 462
Db 367 TYDNDIALQLKSDSSRCAQESSVVRTVCLFPADQLPDWTECELSGYGKHEALSPPYSE 426
QY 463 QLLDAKVKLIANTLCNSRQLYDHMDSDMICAGNLQKPG-----QDTCQDSDGGPLTCEK 517
Db 427 RLKEAHVRLYPSRCTSHLLNRTVTDNMLCAGTRSGGPOANLHDACQDSDGGPLVCLN 486
QY 518 DGTYYVYGVISWGLECGKR--PGVYTVTKFLNWKATIK 555
Db 487 DGRMTLVGIISWGLGCGQKQDVPGYTVKVTNYLDWIRDNR 526

RESULT 8
5185259-8
; Patent No. 5185259
; APPLICANT: GOEDDEL, DAVID V.; KOHR, WILLIAM J.; PENNICA, DIANE;
; VEHAR, GORDON A.
; TITLE OF INVENTION: TRUNCATED HUMAN TISSUE PLASMINOGEN
; ACTIVATOR
; NUMBER OF SEQUENCES: 15
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/489,855
; FILING DATE: 02-MAR-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 12,694
; FILING DATE: 09-FEB-1987
; APPLICATION NUMBER: 483,052
```

```
; FILING DATE: 07-APR-1983
; APPLICATION NUMBER: 398,003
; FILING DATE: 14-JUL-1982
; APPLICATION NUMBER: 374,860
; FILING DATE: 05-MAY-1982
; SEQ ID NO: 8:
; LENGTH: 527
; 5185259-8

Query Match          24.3%; Score 766.5; DB 6; Length 527;
Best Local Similarity 35.0%; Pred. No. 2.1e-48;
Matches 182; Conservative 77; Mismatches 176; Indels 85; Gaps 20;

QY 77 CQNPCEHGGDC--LVHGSTFTCSCLAPFSGNKCO-KVQNTCKDNP--CGRG-----123
Db 51 CSEPRCFNGTCCQALYFSDVFCQCPGFAKCEIDTRATCYEDQGISYRGTWSTAESG 110
QY 124 -QCLITOSPPYRCVCKHPYTGPSCSQVVPVCRPNPCQ-----NGATCSRHKRRSKFTCAC 178
Db 111 AECTNWS-----SALAQKPYSGR-----RPDAIRLGLGNHNYCRNPDNRDSDKPCYV 157
QY 179 PDQFK-GKF-----CEIGSDDCYVGDSYRGKMNRTVNOHACLYNNSHLLQENY 228
Db 158 ---FRAGKYSSEFCSTPACSEGNSDCYFGNGSAYRGTHSLTESGASCLPWNMSMILGKYV 214
QY 229 NMFMEDAETHGIGEHNFERNPDADKPCWFIKVTNDKVKWEYCDVSACSAQDVAYPEESP 288
Db 215 TAQNPQAALGLGKHNYCRNPDGAKPWCHV-LKNRRLTWECYCDVPSCS-----262
QY 289 TEPSTKLPGFDSCKTEIAERKIRYGGFKSTAGKHPWQASLOSPLTITSMPOGHFCG 348
Db 263 -----TCGLRYSQPF-RIKGGLFADIASHPWQAIFAK--HRRSEGERFLCG 308
QY 349 GALIHPCWVLTAACHTD--IKTRHLKVVLGDQDLKKEEFHQSPRVEKIFKYSHYNERDE 406
Db 309 GILLISSCWILSAACHFCQERFPFPHLTVILGRYVVPGEEBQKEVEKYIVHKEFD--DD 366
QY 407 IPHNDIALKLKVPDGHCALESKYVTVCLP--DGSPFSGSECHISGMGVTETGKG--SR 462
Db 367 TYDNDIALQLKSDSSRCAQESSVVRTVCLFPADQLPDWTECELSGYGKHEALSPPYSE 426
QY 463 QLLDAKVKLIANTLCNSRQLYDHMDSDMICAGNLQKPG-----QDTCQDSDGGPLTCEK 517
Db 427 RLKEAHVRLYPSRCTSHLLNRTVTDNMLCAGTRSGGPOANLHDACQDSDGGPLVCLN 486
QY 518 DGTYYVYGVISWGLECGKR--PGVYTVTKFLNWKATIK 555
Db 487 DGRMTLVGIISWGLGCGQKQDVPGYTVKVTNYLDWIRDNR 526

RESULT 9
US-08-811-949-43
; Sequence 43, Application US/08811949
; Patent No. 5840533
; GENERAL INFORMATION:
; APPLICANT: NIWA, MINEO
; APPLICANT: SAITO, YOSHIMASA
; APPLICANT: SASAKI, HITOSHI
; APPLICANT: HAYASHI, MASAKO
; APPLICANT: NOTANI, JOUJI
; APPLICANT: KOBAYASHI, MASAKAZU
; TITLE OF INVENTION: TISSUE PLASMINOGEN ACTIVATOR
; NUMBER OF SEQUENCES: 67
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
; ADDRESSEE: P.C.
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
```

```

; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION NUMBER: US/08/811,949
; FILING DATE: 05-MAR-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: OBLON, NORMAN F.
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 18-966-0
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-413-3000
; TELEFAX: 703-413-2220
; INFORMATION FOR SEQ ID NO: 43:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 562 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-811-949-43

Query Match 24.3%; Score 766.5; DB 2; Length 562;
Best Local Similarity 35.0%; Pred. No. 2.3e-48;
Matches 182; Conservative 77; Mismatches 176; Indels 85; Gaps 20;

QY 77 CQNPCEHGGDC--LVHGSTFTCSCLAPFSGNKQ--KVQNTCKDNP--CGRG-----123
DB 86 CSEPRCFNGTCQQAIFYSDFCVQCPEGFAGKCEIDTRATCYEDQGISYRGTWSTAESG 145
QY 124 -QCLITQSPPYRYRCVKHPYTPGSCSQVVPVCRNPQO----NGATCSRHKRSKFTCAC 178
DB 146 AECTNWS----SALAQKPYSGR-----RPDAIRLGLGNHNYCRNDRSKPWCYV 192
QY 179 PDQFK-GKF-----CEIGSDDCYVGDGYSGYRGKMRVTNQHACLYWNHLLQENY 228
DB 193 ---FKAGKYSSEFCSTPACSEGNDCYFGNGSAVYRGTHSLTESGASCLPWNMILIGKY 249
QY 229 NMFMEDAETHGIGEHNFRCNPDADKPCWCFIKVTNDKVKWEYCDVSACSADQVAYPEBP 288
DB 250 TAQNPQAALGLGKHNYCRNPDGDAKPWCHV-LKNRLTWECYDVPSCS-----297
QY 289 TEPSTKLPGFDSGCKTETAEARKIKIYGGFKSTAGKHPQASLOSLLPLTISMPQGHFCG 348
DB 298 -----TCGLRQVYQPF-RIKGLFADIASHPWQAAIFAK--HRRSPGERFLCG 343
QY 349 GALIHPCWLTAAHCTD--IKTRHLKVLVGDQDLKKEEFHQSFVEKIFKYSHYNERDE 406
DB 344 GILISSCWILSAAHCFQERFPFHLLTVILGRYRVVPGEEQKFEVEKIVHKEFD--DD 401
QY 407 IPHNDIALKLKPVDPDGHCALESKYVKTVCPLP--DGSFPGSGECHISGWGVTETGKG--SR 462
DB 402 TYDNDIALQLKSDSSRCQAQESSVVRTVCLPPADLQLPDWTECELSGVGKHEALSPFYSE 461
QY 463 QLLDAKVKLIANTLNCNSRLYDHMIDDSMICAGNLQKPG-----QDTCCGDSGGPLTCEK 517
DB 462 RLKEAHVRLYPSRSCTSHLLNRTVTDNMLCAGDTRSGGPGQANLHDACQDGGSLVCLN 521
QY 518 DGYVYVGVISWGLECGKR--PGVYTVQVTKFLNWKATIK 555
DB 522 DGRMTLVGLISWGLECGQKDPGVYTKVTNYLDWIRDNR 561

RESULT 10
US-08-560-098A-50
; Sequence 50, Application US/08560098A
; Patent No. 5976841
; GENERAL INFORMATION:
; APPLICANT: WNEBDT, Stephan
; APPLICANT: HEINZEL-WIELAND, Regina
; APPLICANT: STEFFENS, Gerd Josef
; TITLE OF INVENTION: Proteins having Fibrinolytic and
; TITLE OF INVENTION: Coagulation-inhibiting Properties

; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Evenson, McKeown, Edwards & Lenahan
; STREET: 1200 G Street, N.W., Suite 700
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/560,098A
; FILING DATE: 17-NOV-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: P 44 40 892.7
; FILING DATE: 17-NOV-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: EVANS, Joseph D.
; REGISTRATION NUMBER: 26,269
; REFERENCE/DOCKET NUMBER: 148/42448
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 628-8800
; TELEFAX: (202) 628-8844
; INFORMATION FOR SEQ ID NO: 50:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 562 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Protein
US-08-560-098A-50

Query Match 24.3%; Score 766.5; DB 2; Length 562;
Best Local Similarity 35.0%; Pred. No. 2.3e-48;
Matches 182; Conservative 77; Mismatches 176; Indels 85; Gaps 20;

QY 77 CQNPCEHGGDC--LVHGSTFTCSCLAPFSGNKQ--KVQNTCKDNP--CGRG-----123
DB 86 CSEPRCFNGTCQQAIFYSDFCVQCPEGFAGKCEIDTRATCYEDQGISYRGTWSTAESG 145
QY 124 -QCLITQSPPYRYRCVKHPYTPGSCSQVVPVCRNPQO----NGATCSRHKRSKFTCAC 178
DB 146 AECTNWS----SALAQKPYSGR-----RPDAIRLGLGNHNYCRNDRSKPWCYV 192
QY 179 PDQFK-GKF-----CEIGSDDCYVGDGYSGYRGKMRVTNQHACLYWNHLLQENY 228
DB 193 ---FKAGKYSSEFCSTPACSEGNDCYFGNGSAVYRGTHSLTESGASCLPWNMILIGKY 249
QY 229 NMFMEDAETHGIGEHNFRCNPDADKPCWCFIKVTNDKVKWEYCDVSACSADQVAYPEBP 288
DB 250 TAQNPQAALGLGKHNYCRNPDGDAKPWCHV-LKNRLTWECYDVPSCS-----297
QY 289 TEPSTKLPGFDSGCKTETAEARKIKIYGGFKSTAGKHPQASLOSLLPLTISMPQGHFCG 348
DB 298 -----TCGLRQVYQPF-RIKGLFADIASHPWQAAIFAK--HRRSPGERFLCG 343
QY 349 GALIHPCWLTAAHCTD--IKTRHLKVLVGDQDLKKEEFHQSFVEKIFKYSHYNERDE 406
DB 344 GILISSCWILSAAHCFQERFPFHLLTVILGRYRVVPGEEQKFEVEKIVHKEFD--DD 401
QY 407 IPHNDIALKLKPVDPDGHCALESKYVKTVCPLP--DGSFPGSGECHISGWGVTETGKG--SR 462
DB 402 TYDNDIALQLKSDSSRCQAQESSVVRTVCLPPADLQLPDWTECELSGVGKHEALSPFYSE 461
QY 463 QLLDAKVKLIANTLNCNSRLYDHMIDDSMICAGNLQKPG-----QDTCCGDSGGPLTCEK 517
DB 462 RLKEAHVRLYPSRSCTSHLLNRTVTDNMLCAGDTRSGGPGQANLHDACQDGGSLVCLN 521
QY 518 DGYVYVGVISWGLECGKR--PGVYTVQVTKFLNWKATIK 555
DB 522 DGRMTLVGLISWGLECGQKDPGVYTKVTNYLDWIRDNR 561
```

Db 522 DGRMTLVGIISWGLGCGKQVPGVYTKVTNYLDWIRDNR 561

RESULT 11

US-08-883-795A-38
; Sequence 38, Application US/08883795A
; Patent No. 5985607
; GENERAL INFORMATION:
; APPLICANT: Delcuve, Genevieve
; TITLE OF INVENTION: Recombinant DNA Molecules and Expression
; TITLE OF INVENTION: Vectors for Tissue Plasminogen Activator
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BERESKIN & PARR
; STREET: 40 King Street West
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5H 3Y2
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/883,795A
; FILING DATE: 27-JUN-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Gravelle, Micheline
; REGISTRATION NUMBER: 40,261
; REFERENCE/DOCKET NUMBER: 7841-062
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 364-7311
; TELEFAX: (416) 361-1398
; INFORMATION FOR SEQ ID NO: 38:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 562 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein (tpa)
US-08-883-795A-38

Query Match 24.3%; Score 766.5; DB 2; Length 562;
Best Local Similarity 35.0%; Pred. No. 2.3e-48;
Matches 182; Conservative 77; Mismatches 176; Indels 85; Gaps 20;

QY 77 CQNPCEHGDC--LVHGSTFTCSCLAPFSGNKKQ-KVQNTCKDNP--CGRG----- 123
Db 86 CSEPRCFNGTCCQALYFDFVCQCEGAGKCEIDTRATCYEDQGISYRGWTSTAESG 145

QY 124 -OCLITQSPPYRCVKHPTGSCSOVVPVCRPNPCQ----NGATCSRHKRSKFTCAC 178
Db 146 ABCTNWN-----SALAQKPYSGR-----RPDAIRLGLGNHNYCRNPDSDSKPCYV 192

QY 179 PQDFK-GKF-----CEIGSDDCYVGDGYSYRGKMRNTVNHQACLYWNSHLLQENY 228
Db 193 ---FKAGKYSSEFCSTPACSEGNDCYFGNGSAYRGTHSLTESGASCLPWNMILIGKY 249

QY 229 NMFEDAETHGIGEHNFCEPNDAEPWCFIKVTNDKVKWEYCDVSACSAQDVAYPEESP 288
Db 250 TAQNPQAALGLGKHNYCRNPDGDAKPWCHV-LKNRLTWECYDVFSCS----- 297

QY 289 TEPSTKLPGFDSCKTEIAERKIKRIYGGFKSTAGHPQWASLQSSLPITISMPQGHFCG 348
Db 146 ABCTNWN-----SALAQKPYSGR-----RPDAIRLGLGNHNYCRNPDSDSKPCYV 192

QY 179 PQDFK-GKF-----CEIGSDDCYVGDGYSYRGKMRNTVNHQACLYWNSHLLQENY 228
Db 193 ---FKAGKYSSEFCSTPACSEGNDCYFGNGSAYRGTHSLTESGASCLPWNMILIGKY 249

QY 229 NMFEDAETHGIGEHNFCEPNDAEPWCFIKVTNDKVKWEYCDVSACSAQDVAYPEESP 288
Db 250 TAQNPQAALGLGKHNYCRNPDGDAKPWCHV-LKNRLTWECYDVFSCS----- 297

QY 289 TEPSTKLPGFDSCKTEIAERKIKRIYGGFKSTAGHPQWASLQSSLPITISMPQGHFCG 348
Db 298 -----TCGLRQYQPF-RIKGLGFADIASHPWQAIFAK--HRRSPGRFLCG 343

QY 349 GALIHPCWLVTAHCTD--IKTRHLKVLGDQDLKKEEFHEQSFVKEIKFYKSHYNERDE 406
Db 344 GILISSCWILSAACFQERPPPHLTVILGRYVYVPGEEQKFEVEKIVVHKEFD--DD 401

QY 407 IPHNDIALKLKXPVDGHCALESKYKVTYCLP--DGSFPGSGECHISGNGVTTGKG--SR 462
Db 402 TYDNDIALQLQSDSRCAQESSVVRVTCCLPPADLQLPDWTECELSGYGKHEALSPFYSE 461

QY 463 QLLDAKVKLIANTLNSRQLYDHMIDDSMICAGNLQKPG-----ODTCQDGGGGLTCBK 517
Db 462 RLKEAHVRLYPSRCTSQHLLNRTVTDNMLCAGDTRSGGPOANLHDACQDGGGGLVCLN 521

QY 518 DGTYYVYVIGVSWGLECGKG--PGVYTVQVTKFLNWKATIK 555
Db 522 DGRMTLVGIISWGLGCGKQVDFGVYTKVTNYLDWIRDNR 561

RESULT 12

US-09-703-695A-4
; Sequence 4, Application US/09703695A
; Patent No. 6593097
; GENERAL INFORMATION:
; APPLICANT: Xu, Yuan
; TITLE OF INVENTION: REVERSE-PHASE HPLC ASSAY FOR PLASMINOGEN ACTIVATORS
; FILE REFERENCE: F1788R1
; CURRENT APPLICATION NUMBER: US/09/703,695A
; CURRENT FILING DATE: 2000-11-01
; PRIOR APPLICATION NUMBER: US 60/163,607
; PRIOR FILING DATE: 1999-11-04
; NUMBER OF SEQ ID NOS: 4
; SEQ ID NO 4
; LENGTH: 562
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-703-695A-4

Query Match 24.3%; Score 766.5; DB 4; Length 562;
Best Local Similarity 35.0%; Pred. No. 2.3e-48;
Matches 182; Conservative 77; Mismatches 176; Indels 85; Gaps 20;

QY 77 CQNPCEHGDC--LVHGSTFTCSCLAPFSGNKKQ-KVQNTCKDNP--CGRG----- 123
Db 86 CSEPRCFNGTCCQALYFDFVCQCEGAGKCEIDTRATCYEDQGISYRGWTSTAESG 145

QY 124 -OCLITQSPPYRCVKHPTGSCSOVVPVCRPNPCQ----NGATCSRHKRSKFTCAC 178
Db 146 ABCTNWN-----SALAQKPYSGR-----RPDAIRLGLGNHNYCRNPDSDSKPCYV 192

QY 179 PQDFK-GKF-----CEIGSDDCYVGDGYSYRGKMRNTVNHQACLYWNSHLLQENY 228
Db 193 ---FKAGKYSSEFCSTPACSEGNDCYFGNGSAYRGTHSLTESGASCLPWNMILIGKY 249

QY 229 NMFEDAETHGIGEHNFCEPNDAEPWCFIKVTNDKVKWEYCDVSACSAQDVAYPEESP 288
Db 250 TAQNPQAALGLGKHNYCRNPDGDAKPWCHV-LKNRLTWECYDVFSCS----- 297

QY 289 TEPSTKLPGFDSCKTEIAERKIKRIYGGFKSTAGHPQWASLQSSLPITISMPQGHFCG 348
Db 298 -----TCGLRQYQPF-RIKGLGFADIASHPWQAIFAK--HRRSPGRFLCG 343

QY 349 GALIHPCWLVTAHCTD--IKTRHLKVLGDQDLKKEEFHEQSFVKEIKFYKSHYNERDE 406
Db 344 GILISSCWILSAACFQERPPPHLTVILGRYVYVPGEEQKFEVEKIVVHKEFD--DD 401

QY 407 IPHNDIALKLKXPVDGHCALESKYKVTYCLP--DGSFPGSGECHISGNGVTTGKG--SR 462
Db 402 TYDNDIALQLKSDSRCAQESSVVRVTCCLPPADLQLPDWTECELSGYGKHEALSPFYSE 461

QY 463 QLLDAKVKLIANTLNSRQLYDHMIDDSMICAGNLQKPG-----ODTCQDGGGGLTCBK 517
Db 462 RLKEAHVRLYPSRCTSQHLLNRTVTDNMLCAGDTRSGGPOANLHDACQDGGGGLVCLN 521

QY 518 DGTYYVYVIGVSWGLECGKR--PGVYTVQVTKFLNWKATIK 555
Db 522 DGRMTLVGIISWGLGCGKQVDFGVYTKVTNYLDWIRDNR 561

RESULT 13
5185259-3
; Patent No. 5185259
; APPLICANT: GOEDDEL, DAVID V.; KOHR, WILLIAM J.; PENNICA, DIANE;
; VEHAR, GORDON A.
; TITLE OF INVENTION: TRUNCATED HUMAN TISSUE PLASMINOGEN
; ACTIVATOR
; NUMBER OF SEQUENCES: 15
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/489,855
; FILING DATE: 02-MAR-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 12,694
; FILING DATE: 09-FEB-1987
; APPLICATION NUMBER: 483,052
; FILING DATE: 07-APR-1983
; APPLICATION NUMBER: 398,003
; FILING DATE: 14-JUL-1982
; APPLICATION NUMBER: 374,860
; FILING DATE: 05-MAY-1982
; SEQ ID NO: 3
; LENGTH: 562
5185259-3

Query Match 24.3%; Score 766.5; DB 6; Length 562;
Best Local Similarity 35.0%; Pred. No. 2.3e-48;
Matches 182; Conservative 77; Mismatches 176; Indels 85; Gaps 20;
QY 77 CQNPCEHGGDC--LVHGSTFTCSCLAPFSGNKQ--KVQNTCKDNP--CGRG-----123
Db 86 CSEPRCENGTCQQAALYFSDVFCQPEGFAKCCIEIDTRATCYEDQGISYRGWTWSTAESG 145
QY 124 -QCLITQSPPYRYRCVKHPYTGPCSCQVVPVCRNPQ-----NGATCSRHKRSKFTCAC 178
Db 146 AECTNWS-----SALAQKPYSGR-----RDAIRLGLGNHNYCRNPDROSKPWCYV 192
QY 179 PDQFK-GKF-----CEIGSDDCYVGDGYSYRGKXNRTVNOHACLYNWSHLLIQENY 228
Db 193 ---FKAGKYSSEFCSTPACSEGNSDCYFGNGSAYRGTHSLTESGASCLPWNMILIGKVY 249
QY 229 NMFMEDAETHGIGHNFCRNPDADEKPCWCFIKVNDKVMWEYCDVVSACSADVAYPEESP 288
Db 250 TAQNPQAALGLGKXNYCRNPDGDAKPCWCHV-LKNRLTWECYCDVPCS-----297
QY 289 TEPSTKLPFCPSGCKTEIAERKIKRIYGGFKSTAGKHPQWASLOSSPLTISMPOGHFCG 348
Db 298 -----TCGLRQYSQPF-RIKGGLFADIASHPWQAIFAK--HRRSPGERFLCG 343
QY 349 GALIHPCWVLTAAHCTD--IKTRHLKVVLGQDLKKEEFHQSPRVEKIFKYSHYNERDE 406
Db 344 GILISSCWILSAAHCFQERFPFHLTVILGRTYRVVPGEEBQKEVEKIVHKEFD--DD 401
QY 407 IPHNDIALLKVPDGHCALESKYVTVCLP--DGSPPSGSECHISGMVGTETGK--SR 462
Db 402 TYDNDIALQLKSDSSRCAQESSVVRTVCLPPADLQLPDWTECELSGYGHEALSPPYSE 461
QY 463 QLLDAKVKLIANTLNCNSQLYDHMIDDSMI CAGNLQKPG-----QDTCCGDSGGPLTCEK 517
Db 462 RLKEAHVRLYSSRSTCSQHLNRTVTDNMLCAGDTRSGGPOANLHDACQSGSGGPLVCLN 521
QY 518 DGTYYVYGVISVGLGCGKR--PGVTVQTKFLNMIKATIK 555
Db 522 DGRMTLVGIISGLGCGQKDPGVYTKVTNYLDWIRDNR 561

RESULT 14
5200340-2
; Patent No. 5200340
; APPLICANT: FOSTER, DONALD C.; MULVIHILL, EILEEN R.; O'HARA,
; PATRICK J.; PINGEL, KURT; YOSHITAKE, SHINJI
; TITLE OF INVENTION: THROMBIN-ACTIVATED TISSUE PLASMINOGEN
; ACTIVATORS
; NUMBER OF SEQUENCES: 34
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/489,855
; FILING DATE: 02-MAR-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 12,694
; FILING DATE: 09-FEB-1987
; APPLICATION NUMBER: 483,052
; FILING DATE: 07-APR-1983
; APPLICATION NUMBER: 398,003
; FILING DATE: 14-JUL-1982
; APPLICATION NUMBER: 374,860
; FILING DATE: 05-MAY-1982
; SEQ ID NO: 3
; LENGTH: 562
5200340-2

Query Match 24.3%; Score 766.5; DB 6; Length 562;
Best Local Similarity 35.0%; Pred. No. 2.3e-48;
Matches 182; Conservative 77; Mismatches 176; Indels 85; Gaps 20;
QY 77 CQNPCEHGGDC--LVHGSTFTCSCLAPFSGNKQ--KVQNTCKDNP--CGRG-----123
Db 86 CSEPRCENGTCQQAALYFSDVFCQPEGFAKCCIEIDTRATCYEDQGISYRGWTWSTAESG 145
QY 124 -QCLITQSPPYRYRCVKHPYTGPCSCQVVPVCRNPQ-----NGATCSRHKRSKFTCAC 178
Db 146 AECTNWS-----SALAQKPYSGR-----RDAIRLGLGNHNYCRNPDROSKPWCYV 192
QY 179 PDQFK-GKF-----CEIGSDDCYVGDGYSYRGKXNRTVNOHACLYNWSHLLIQENY 228
Db 193 ---FKAGKYSSEFCSTPACSEGNSDCYFGNGSAYRGTHSLTESGASCLPWNMILIGKVY 249
QY 229 NMFMEDAETHGIGHNFCRNPDADEKPCWCFIKVNDKVMWEYCDVVSACSADVAYPEESP 288
Db 250 TAQNPQAALGLGKXNYCRNPDGDAKPCWCHV-LKNRLTWECYCDVPCS-----297
QY 289 TEPSTKLPFCPSGCKTEIAERKIKRIYGGFKSTAGKHPQWASLOSSPLTISMPOGHFCG 348
Db 298 -----TCGLRQYSQPF-RIKGGLFADIASHPWQAIFAK--HRRSPGERFLCG 343
QY 349 GALIHPCWVLTAAHCTD--IKTRHLKVVLGQDLKKEEFHQSPRVEKIFKYSHYNERDE 406
Db 344 GILISSCWILSAAHCFQERFPFHLTVILGRTYRVVPGEEBQKEVEKIVHKEFD--DD 401
QY 407 IPHNDIALLKVPDGHCALESKYVTVCLP--DGSPPSGSECHISGMVGTETGK--SR 462
Db 402 TYDNDIALQLKSDSSRCAQESSVVRTVCLPPADLQLPDWTECELSGYGHEALSPPYSE 461
QY 463 QLLDAKVKLIANTLNCNSQLYDHMIDDSMI CAGNLQKPG-----QDTCCGDSGGPLTCEK 517
Db 462 RLKEAHVRLYSSRSTCSQHLNRTVTDNMLCAGDTRSGGPOANLHDACQSGSGGPLVCLN 521
QY 518 DGTYYVYGVISVGLGCGKR--PGVTVQTKFLNMIKATIK 555
Db 522 DGRMTLVGIISGLGCGQKDPGVYTKVTNYLDWIRDNR 561

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/53,412
; FILING DATE: 22-MAY-1987
; SEQ ID NO: 2
; LENGTH: 562
5200340-2

Query Match 24.3%; Score 766.5; DB 6; Length 562;
Best Local Similarity 35.0%; Pred. No. 2.3e-48;
Matches 182; Conservative 77; Mismatches 176; Indels 85; Gaps 20;
QY 77 CQNPCEHGGDC--LVHGSTFTCSCLAPFSGNKQ--KVQNTCKDNP--CGRG-----123
Db 86 CSEPRCENGTCQQAALYFSDVFCQPEGFAKCCIEIDTRATCYEDQGISYRGWTWSTAESG 145
QY 124 -QCLITQSPPYRYRCVKHPYTGPCSCQVVPVCRNPQ-----NGATCSRHKRSKFTCAC 178
Db 146 AECTNWS-----SALAQKPYSGR-----RDAIRLGLGNHNYCRNPDROSKPWCYV 192
QY 179 PDQFK-GKF-----CEIGSDDCYVGDGYSYRGKXNRTVNOHACLYNWSHLLIQENY 228
Db 193 ---FKAGKYSSEFCSTPACSEGNSDCYFGNGSAYRGTHSLTESGASCLPWNMILIGKVY 249
QY 229 NMFMEDAETHGIGHNFCRNPDADEKPCWCFIKVNDKVMWEYCDVVSACSADVAYPEESP 288
Db 250 TAQNPQAALGLGKXNYCRNPDGDAKPCWCHV-LKNRLTWECYCDVPCS-----297
QY 289 TEPSTKLPFCPSGCKTEIAERKIKRIYGGFKSTAGKHPQWASLOSSPLTISMPOGHFCG 348
Db 298 -----TCGLRQYSQPF-RIKGGLFADIASHPWQAIFAK--HRRSPGERFLCG 343
QY 349 GALIHPCWVLTAAHCTD--IKTRHLKVVLGQDLKKEEFHQSPRVEKIFKYSHYNERDE 406
Db 344 GILISSCWILSAAHCFQERFPFHLTVILGRTYRVVPGEEBQKEVEKIVHKEFD--DD 401
QY 407 IPHNDIALLKVPDGHCALESKYVTVCLP--DGSPPSGSECHISGMVGTETGK--SR 462
Db 402 TYDNDIALQLKSDSSRCAQESSVVRTVCLPPADLQLPDWTECELSGYGHEALSPPYSE 461
QY 463 QLLDAKVKLIANTLNCNSQLYDHMIDDSMI CAGNLQKPG-----QDTCCGDSGGPLTCEK 517
Db 462 RLKEAHVRLYSSRSTCSQHLNRTVTDNMLCAGDTRSGGPOANLHDACQSGSGGPLVCLN 521
QY 518 DGTYYVYGVISVGLGCGKR--PGVTVQTKFLNMIKATIK 555
Db 522 DGRMTLVGIISGLGCGQKDPGVYTKVTNYLDWIRDNR 561

RESULT 15
5344773-2
; Patent No. 5344773
; APPLICANT: WEI, CHA-MER; HSIUNG, NANCY; REDDY, VERMURI B.;
; LEMONT, JEFFREY F.; DACKOWSKI, WILLIAM; DOUGLAS, RICHARD;
; COLE, EDWARD S.; PURCELL JR., RICHARD D.; LAU, DAVID TAI-YUI
; TITLE OF INVENTION: HUMAN UTERINE TISSUE PLASMINOGEN
; ACTIVATOR PRODUCED BY RECOMBIANT DNA
; NUMBER OF SEQUENCES: 6
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/782,686
; FILING DATE: 01-OCT-1985
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 656,770
; FILING DATE: 01-OCT-1984
; SEQ ID NO: 2
; LENGTH: 562
5344773-2

Query Match 24.3%; Score 766.5; DB 6; Length 562;
Best Local Similarity 35.0%; Pred. No. 2.3e-48;
Matches 182; Conservative 77; Mismatches 176; Indels 85; Gaps 20;
QY 77 CQNPCEHGGDC--LVHGSTFTCSCLAPFSGNKQ--KVQNTCKDNP--CGRG-----123
Db 86 CSEPRCENGTCQQAALYFSDVFCQPEGFAKCCIEIDTRATCYEDQGISYRGWTWSTAESG 145
QY 124 -QCLITQSPPYRYRCVKHPYTGPCSCQVVPVCRNPQ-----NGATCSRHKRSKFTCAC 178
Db 146 AECTNWS-----SALAQKPYSGR-----RDAIRLGLGNHNYCRNPDROSKPWCYV 192
QY 179 PDQFK-GKF-----CEIGSDDCYVGDGYSYRGKXNRTVNOHACLYNWSHLLIQENY 228
Db 193 ---FKAGKYSSEFCSTPACSEGNSDCYFGNGSAYRGTHSLTESGASCLPWNMILIGKVY 249
QY 229 NMFMEDAETHGIGHNFCRNPDADEKPCWCFIKVNDKVMWEYCDVVSACSADVAYPEESP 288
Db 250 TAQNPQAALGLGKXNYCRNPDGDAKPCWCHV-LKNRLTWECYCDVPCS-----297
QY 289 TEPSTKLPFCPSGCKTEIAERKIKRIYGGFKSTAGKHPQWASLOSSPLTISMPOGHFCG 348
Db 298 -----TCGLRQYSQPF-RIKGGLFADIASHPWQAIFAK--HRRSPGERFLCG 343
QY 349 GALIHPCWVLTAAHCTD--IKTRHLKVVLGQDLKKEEFHQSPRVEKIFKYSHYNERDE 406
Db 344 GILISSCWILSAAHCFQERFPFHLTVILGRTYRVVPGEEBQKEVEKIVHKEFD--DD 401
QY 407 IPHNDIALLKVPDGHCALESKYVTVCLP--DGSPPSGSECHISGMVGTETGK--SR 462
Db 402 TYDNDIALQLKSDSSRCAQESSVVRTVCLPPADLQLPDWTECELSGYGHEALSPPYSE 461
QY 463 QLLDAKVKLIANTLNCNSQLYDHMIDDSMI CAGNLQKPG-----QDTCCGDSGGPLTCEK 517
Db 462 RLKEAHVRLYSSRSTCSQHLNRTVTDNMLCAGDTRSGGPOANLHDACQSGSGGPLVCLN 521
QY 518 DGTYYVYGVISVGLGCGKR--PGVTVQTKFLNMIKATIK 555
Db 522 DGRMTLVGIISGLGCGQKDPGVYTKVTNYLDWIRDNR 561

Db 86 CSEPRCNGGTCQALYFDFVQCPEGFAGKCEIDTRATCYEDQGISYRGTWSTAESG 145
QY 124 -OCLITOSPYYRCVCKHPYTGPCSQVPCRENPCQ---NGATCSRHKRBSKFTCAC 178
Db 146 AECTNWN---SALAOKPYSGR-----RDAIRLGLGNHNYCNDRDSKPCYV 192
QY 179 PDQFK-GXF-----CEIGSDDCYVGDGYSGKXNRTVNOHACLYWNSHLLQENY 228
Db 193 ---FKAGKYSSEFCSTPACSEGNDCYFGNGSAYRGTHSLTESGASCLPWNSMILIGKY 249
QY 229 NMFMEDAETHGIGBNFCENPDADKPKCFIKVTNDKVKWEYCDVSACSQAQDVAYPEEP 288
Db 250 TAQNPSAQLGLGKHNYCRNPDGAKPWCHV-LKNRRLTWIYCDVPSCS----- 297
QY 289 TEPSTKLPFGDFSCGKTEIAERKIKRIYGGFKSTAGKHPWQASLOSLELTISMPCQHFCG 348
Db 298 -----TCGLRQYSQPF-RKGGLPADIASHPWQAAIFAK-HRRSPGERFLCG 343
QY 349 GALIHPCWLTAAHCTD--IKTRHLKVVLGDQDLKSEFHEQSPRVEKIFKYSHYNERDE 406
Db 344 GILISSCWILSAAHCFQERFPFPHLTVILGRTRYVVPGEESQKPEVEKYIVHKEFD--DD 401
QY 407 IPHNDIALLLKLPVDGHCALSXYKTVCLP--DGSPFGSGECHISGWGTETGKG--SR 462
Db 402 TYDNDIALLLKSDSSRCAQSSVWRTVCLPADLQLPDWTCELSGVGKHEALS PFYSE 461
QY 463 QLLDAKYKLANTLNSRQLYDHMIDDSMICAGNLQKEG-----QPTCGDSGGPLTCEK 517
Db 462 RLKEAHVRLYPSSRCTSQHLLNRTVTDNMLCAGDTRSGGPOANLHDACQDSGGPLVCLN 521
QY 518 DGTYYVYGVISWGLECGKR--PGVYTVTKFLNWKATIK 555
Db 522 DGRMTLVGLIISWGLCGGQKXDPGVYTKVTNYLDWIRDNR 561

Search completed: May 24, 2004, 09:50:30
Job time : 19.5 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 24, 2004, 09:49:06 ; Search time 42 Seconds
(without alignments)
3718.956 Million cell updates/sec

Title: US-09-912-559-3
Perfect score: 3154
Sequence: 1 MFASMSDLHVLLMALVGTKT.....TQVTKFLNWKATIKSES GF 560

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1149313 seqs, 278921704 residues

Total number of hits satisfying chosen parameters: 1149313

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA:
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/FCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	3154	100.0	560	9	US-09-912-559-3
2	3154	100.0	560	14	US-10-172-712-32
3	3154	100.0	560	15	US-10-391-215-5
4	3151	99.9	560	15	US-10-391-215-7
5	3146	99.7	560	15	US-10-391-215-6
6	3143	99.7	560	9	US-09-912-559-4
7	3143	99.7	560	15	US-10-391-215-8
8	832.5	26.4	555	14	US-10-172-712-28
9	767.5	24.3	562	9	US-09-974-298-145
10	766.5	24.3	527	10	US-09-987-457-18
11	766.5	24.3	527	10	US-09-987-455-19
12	766.5	24.3	527	12	US-10-432-842-1
13	766.5	24.3	527	15	US-10-360-101-203
14	766.5	24.3	562	9	US-09-969-271-7
15	766.5	24.3	562	12	US-10-411-037-26

Sequence 26, Appl
Sequence 8, Appl
Sequence 4, Appl
Sequence 26, Appl
Sequence 26, Appl
Sequence 1, Appl
Sequence 2, Appl
Sequence 30, Appl
Sequence 3, Appl
Sequence 3, Appl
Sequence 10, Appl
Sequence 11, Appl
Sequence 8, Appl
Sequence 591, App
Sequence 2, Appl
Sequence 3, Appl
Sequence 34, Appl
Sequence 34, Appl
Sequence 2, Appl
Sequence 184, App
Sequence 161, App
Sequence 21, Appl
Sequence 414, App
Sequence 1275, App
Sequence 34, Appl
Sequence 594, App
Sequence 149, App
Sequence 71, Appl

562 12 US-10-411-026-26
562 14 US-10-193-656-8
562 14 US-10-443-701-4
562 16 US-10-410-962-26
562 16 US-10-411-049-26
562 15 US-10-401-077-1
650 15 US-09-858-909-2
615 14 US-10-172-712-30
615 14 US-10-172-712-30
615 16 US-10-449-132-2
372 9 US-09-084-491A-3
372 13 US-10-103-704-3
354 10 US-09-987-457-10
354 10 US-09-987-455-11
377 10 US-09-987-455-8
433 12 US-10-087-192-591
411 15 US-10-407-821-2
411 9 US-09-880-503-3
431 12 US-10-411-037-34
431 12 US-10-411-026-34
431 13 US-10-076-421-2
431 14 US-10-171-311-184
431 14 US-10-301-822-161
431 14 US-10-131-985-21
431 15 US-10-295-027-414
431 15 US-10-295-027-1275
431 16 US-10-410-962-34
431 16 US-10-411-049-34
437 12 US-10-087-192-594
431 14 US-10-247-671-149
812 12 US-09-825-751A-71

ALIGNMENTS

RESULT 1

US-09-912-559-3
; Sequence 3, Application US/09912559
; Patent No. US20020142316A1
; GENERAL INFORMATION:
; APPLICANT: ROEMISCH, JURGEN
; APPLICANT: STOEHR, HANS-ARNOLD
; APPLICANT: FEUSNER, ANNETTE
; APPLICANT: LANG WIEGAND
; APPLICANT: WEIMER, THOMAS
; APPLICANT: BECKER, MARGRET
; APPLICANT: NERLICH, CLAUDIA
; APPLICANT: MUTH-NAUMANN, GUDRUN
; TITLE OF INVENTION: MUTANTS OF THE FACTOR VII-ACTIVATING PROTEASE AND
; TITLE OF INVENTION: DETECTION METHODS USING SPECIFIC ANTIBODIES
; FILE REFERENCE: 06478.1457
; CURRENT APPLICATION NUMBER: US/09/912,559
; CURRENT FILING DATE: 2001-07-26
; PRIOR APPLICATION NUMBER: DE 100 36 641.4
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: DE 100 50 040.4
; PRIOR FILING DATE: 2000-10-10
; PRIOR APPLICATION NUMBER: DE 100 52 319.6
; PRIOR FILING DATE: 2000-10-21
; PRIOR APPLICATION NUMBER: DE 101 18 706.8
; PRIOR FILING DATE: 2001-04-12
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 560
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-912-559-3

Query Match 100.0%; Score 3154; DB 9; Length 560;
Best Local Similarity 100.0%; Pred. No. 9.1e-241;
Matches 560; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MFARMSDLHVLLMALVGTACGFSLSLLESIDPDWTPDQYDYSYEDYNOENTSSLT 60
Db 1 MFARMSDLHVLLMALVGTACGFSLSLLESIDPDWTPDQYDYSYEDYNOENTSSLT 60
QY 61 HAENPDWYTEDQADPCQPNPCFCEHGGDCLVHGSTFTTCSCLAPPSGNKCKQVQNTCKDNPC 120
Db 61 HAENPDWYTEDQADPCQPNPCFCEHGGDCLVHGSTFTTCSCLAPPSGNKCKQVQNTCKDNPC 120
QY 121 GRGQCLITQSPPPYRCVCKHPYTGPCSQVVPVCRPNPCQNGATCSRHKERSKFTCACPD 180
Db 121 GRGQCLITQSPPPYRCVCKHPYTGPCSQVVPVCRPNPCQNGATCSRHKERSKFTCACPD 180
QY 181 QFKGKFCIEGSDDCYVGDGYSYRGKXNRTVNHACLYNNSHLLQENYNNFMEDAETHGI 240
Db 181 QFKGKFCIEGSDDCYVGDGYSYRGKXNRTVNHACLYNNSHLLQENYNNFMEDAETHGI 240
QY 241 GEHNFENPNDADEKPCWCFIKVTNDKWKVEYCDVSASQAQVAYPEESPTSTKLPGFDS 300
Db 241 GEHNFENPNDADEKPCWCFIKVTNDKWKVEYCDVSASQAQVAYPEESPTSTKLPGFDS 300
QY 301 CGKTEIAERIKRIYGGFKSTAGKHPQASLQSSPLTISMPQGHFCGGALIHPCWVLTA 360
Db 301 CGKTEIAERIKRIYGGFKSTAGKHPQASLQSSPLTISMPQGHFCGGALIHPCWVLTA 360
QY 361 AHCTDIKTRHLKVVLGDQDLKKEEFHQSPRVEKIFKYSHYNERDEIPHNDIALLKLPV 420
Db 361 AHCTDIKTRHLKVVLGDQDLKKEEFHQSPRVEKIFKYSHYNERDEIPHNDIALLKLPV 420
QY 421 DGHCALESKYVKTVCPLDGGFPSPGSECHISGWGTETGKSRQLLDAKVLIANTLCNSR 480
Db 421 DGHCALESKYVKTVCPLDGGFPSPGSECHISGWGTETGKSRQLLDAKVLIANTLCNSR 480
QY 481 QLYDHMIDDSMICAGNLQKPGQDTCCQDSGGPLTCEKDGYYVYVGIWSGLECGKRPVY 540
Db 481 QLYDHMIDDSMICAGNLQKPGQDTCCQDSGGPLTCEKDGYYVYVGIWSGLECGKRPVY 540
QY 541 TQVTKFLNWKATIKSESSEGF 560
Db 541 TQVTKFLNWKATIKSESSEGF 560

RESULT 2
US-10-172-712-32
; Sequence 32, Application US/10172712
; Publication No. US20030125232A1
; GENERAL INFORMATION:
; APPLICANT: GRIFFIN, JOHN H.
; APPLICANT: GALE, ANDREW J.
; APPLICANT: PELLEOUE, JEAN-LUC
; APPLICANT: GETZOFF, ELIZABETH D.
; TITLE OF INVENTION: STABILIZED PROTEINS WITH ENGINEERED DISULFIDE BONDS
; FILE REFERENCE: 4198-4001US1
; CURRENT APPLICATION NUMBER: US/10/172,712
; PRIOR FILING DATE: 2002-09-30
; PRIOR APPLICATION NUMBER: 60/298,578
; PRIOR FILING DATE: 2001-06-14
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 32
; LENGTH: 560
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-172-712-32

Query Match 100.0%; Score 3154; DB 14; Length 560;
Best Local Similarity 100.0%; Pred. No. 9.1e-241;
Matches 560; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MFARMSDLHVLLMALVGTACGFSLSLLESIDPDWTPDQYDYSYEDYNOENTSSLT 60
Db 1 MFARMSDLHVLLMALVGTACGFSLSLLESIDPDWTPDQYDYSYEDYNOENTSSLT 60
QY 61 HAENPDWYTEDQADPCQPNPCFCEHGGDCLVHGSTFTTCSCLAPPSGNKCKQVQNTCKDNPC 120

Db 61 HAENPDWYTEDQADPCQPNPCFCEHGGDCLVHGSTFTTCSCLAPPSGNKCKQVQNTCKDNPC 120
QY 121 GRGQCLITQSPPPYRCVCKHPYTGPCSQVVPVCRPNPCQNGATCSRHKERSKFTCACPD 180
Db 121 GRGQCLITQSPPPYRCVCKHPYTGPCSQVVPVCRPNPCQNGATCSRHKERSKFTCACPD 180
QY 181 QFKGKFCIEGSDDCYVGDGYSYRGKXNRTVNHACLYNNSHLLQENYNNFMEDAETHGI 240
Db 181 QFKGKFCIEGSDDCYVGDGYSYRGKXNRTVNHACLYNNSHLLQENYNNFMEDAETHGI 240
QY 241 GEHNFENPNDADEKPCWCFIKVTNDKWKVEYCDVSASQAQVAYPEESPTSTKLPGFDS 300
Db 241 GEHNFENPNDADEKPCWCFIKVTNDKWKVEYCDVSASQAQVAYPEESPTSTKLPGFDS 300
QY 301 CGKTEIAERIKRIYGGFKSTAGKHPQASLQSSPLTISMPQGHFCGGALIHPCWVLTA 360
Db 301 CGKTEIAERIKRIYGGFKSTAGKHPQASLQSSPLTISMPQGHFCGGALIHPCWVLTA 360
QY 361 AHCTDIKTRHLKVVLGDQDLKKEEFHQSPRVEKIFKYSHYNERDEIPHNDIALLKLPV 420
Db 361 AHCTDIKTRHLKVVLGDQDLKKEEFHQSPRVEKIFKYSHYNERDEIPHNDIALLKLPV 420
QY 421 DGHCALESKYVKTVCPLDGGFPSPGSECHISGWGTETGKSRQLLDAKVLIANTLCNSR 480
Db 421 DGHCALESKYVKTVCPLDGGFPSPGSECHISGWGTETGKSRQLLDAKVLIANTLCNSR 480
QY 481 QLYDHMIDDSMICAGNLQKPGQDTCCQDSGGPLTCEKDGYYVYVGIWSGLECGKRPVY 540
Db 481 QLYDHMIDDSMICAGNLQKPGQDTCCQDSGGPLTCEKDGYYVYVGIWSGLECGKRPVY 540
QY 541 TQVTKFLNWKATIKSESSEGF 560
Db 541 TQVTKFLNWKATIKSESSEGF 560

RESULT 3
US-10-391-215-5
; Sequence 5, Application US/10391215
; Publication No. US20040009543A1
; GENERAL INFORMATION:
; APPLICANT: KIECHL, STEFAN
; APPLICANT: WILLEIT, JOHANN
; APPLICANT: WIEDERMANN, CHRISTIAN JOSEF
; APPLICANT: ROEMISCH, JUERGEN
; APPLICANT: WEIMER, THOMAS
; APPLICANT: FEUSSNER, ANNETTE
; APPLICANT: STOEHR, HANS-ARNOLD
; APPLICANT: DOERSAM, VOLKER
; TITLE OF INVENTION: MARBURG I MUTANT OF FACTOR VII ACTIVATING PROTEASE
; FILE REFERENCE: 06478.1457-01
; CURRENT APPLICATION NUMBER: US/10/391,215
; PRIOR FILING DATE: 2003-03-19
; PRIOR APPLICATION NUMBER: 09/912,559
; PRIOR FILING DATE: 2001-07-26
; PRIOR APPLICATION NUMBER: DE 100 36 641.4
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: DE 100 50 040.4
; PRIOR FILING DATE: 2000-10-10
; PRIOR APPLICATION NUMBER: DE 100 52 319.6
; PRIOR FILING DATE: 2000-10-21
; PRIOR APPLICATION NUMBER: DE 101 18 706.8
; PRIOR FILING DATE: 2001-04-12
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 560
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-391-215-5

Query Match 100.0%; Score 3154; DB 15; Length 560;

Best Local Similarity 100.0%; Pred. No. 9.1e-241; Matches 560; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MFARMSDLHVLMLALVGTACGFSLSLLESLEDPDWTDPQDYSDYEDYNQENNTSSTLT 60

Db 1 MFARMSDLHVLMLALVGTACGFSLSLLESLEDPDWTDPQDYSDYEDYNQENNTSSTLT 60

QY 61 HAENPDWYTTEDQADPCQNPCEHGGDCLVHGSTFTCSCLAPFSGNKCKQVQNTCKDNPC 120

Db 61 HAENPDWYTTEDQADPCQNPCEHGGDCLVHGSTFTCSCLAPFSGNKCKQVQNTCKDNPC 120

QY 121 GRGQCLITQSPPPYRVCVKHPYTGPSCSQVVPVCRPNPCQNGATCSRHKRSKFTACPD 180

Db 121 GRGQCLITQSPPPYRVCVKHPYTGPSCSQVVPVCRPNPCQNGATCSRHKRSKFTACPD 180

QY 181 QFKGKFCBEGSDDCYVGDGYSGRGNRTVNHACLYWNSHLLLOENYMFMEDEATHGI 240

Db 181 QFKGKFCBEGSDDCYVGDGYSGRGNRTVNHACLYWNSHLLLOENYMFMEDEATHGI 240

QY 241 GEHNFCRNPDADKPCWCFIKVTNDKVWEYCDVSACSADQVAYPEESPTSTKLPGFDS 300

Db 241 GEHNFCRNPDADKPCWCFIKVTNDKVWEYCDVSACSADQVAYPEESPTSTKLPGFDS 300

QY 301 CGKTEIAERKIKRIYGGFKSTAGKHPWQASLOSSLPLTISMPQGHFCGGLIHPCWVLTA 360

Db 301 CGKTEIAERKIKRIYGGFKSTAGKHPWQASLOSSLPLTISMPQGHFCGGLIHPCWVLTA 360

QY 361 AHCTDIKTRHLKVVLGDQDLKKEEFHQSFVEKIFKYSHYNERDEIPHNDIALLLKPV 420

Db 361 AHCTDIKTRHLKVVLGDQDLKKEEFHQSFVEKIFKYSHYNERDEIPHNDIALLLKPV 420

QY 421 DGHCALESKYVKTVCCLPDGSPSGSECHISGMWVETGKGRQLLDKAKVLIANTLNSR 480

Db 421 DGHCALESKYVKTVCCLPDGSPSGSECHISGMWVETGKGRQLLDKAKVLIANTLNSR 480

QY 481 QLYDHMIDDSMICAGNLQKPGQDTCEKDGPTCEKDGTYVYVGVISWGLECGKRPVY 540

Db 481 QLYDHMIDDSMICAGNLQKPGQDTCEKDGPTCEKDGTYVYVGVISWGLECGKRPVY 540

QY 541 TQVTKFLNWKATIKSESSEGF 560

Db 541 TQVTKFLNWKATIKSESSEGF 560

RESULT 4

US-10-391-215-7

; Sequence 7, Application US/10391215

; Publication No. US20040009543A1

; GENERAL INFORMATION:

; APPLICANT: KIECHL, STEFAN

; APPLICANT: WILLEIT, JOHANN

; APPLICANT: WIEDERMANN, CHRISTIAN JOSEF

; APPLICANT: ROEMISCH, JUERGEN

; APPLICANT: WEIMER, THOMAS

; APPLICANT: FEUSSNER, ANNETTE

; APPLICANT: STOEHR, HANS-ARNOLD

; APPLICANT: DOERSAM, VOLKER

; TITLE OF INVENTION: MARBURG I MUTANT OF FACTOR VII ACTIVATING PROTEASE

; TITLE OF INVENTION: (FSAP) AS RISK FACTOR FOR ARTERIAL THROMBOSIS

; FILE REFERENCE: 06478.1457-01

; CURRENT FILING DATE: 2003-03-19

; CURRENT APPLICATION NUMBER: US/10/391.215

; PRIOR FILING DATE: 2001-07-26

; PRIOR FILING DATE: 2000-07-26

; PRIOR FILING DATE: 2000-10-10

; PRIOR FILING DATE: 2000-10-10

; PRIOR FILING DATE: 2000-10-21

; PRIOR FILING DATE: 2001-04-12

; NUMBER OF SEQ ID NOS: 8

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 7

; LENGTH: 560

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-391-215-7

Query Match 99.9%; Score 3151; DB 15; Length 560;

Best Local Similarity 99.8%; Pred. No. 1.6e-240; Indels 0; Gaps 0;

Matches 559; Conservative 1; Mismatches 0;

QY 1 MFARMSDLHVLMLALVGTACGFSLSLLESLEDPDWTDPQDYSDYEDYNQENNTSSTLT 60

Db 1 MFARMSDLHVLMLALVGTACGFSLSLLESLEDPDWTDPQDYSDYEDYNQENNTSSTLT 60

QY 61 HAENPDWYTTEDQADPCQNPCEHGGDCLVHGSTFTCSCLAPFSGNKCKQVQNTCKDNPC 120

Db 61 HAENPDWYTTEDQADPCQNPCEHGGDCLVHGSTFTCSCLAPFSGNKCKQVQNTCKDNPC 120

QY 121 GRGQCLITQSPPPYRVCVKHPYTGPSCSQVVPVCRPNPCQNGATCSRHKRSKFTACPD 180

Db 121 GRGQCLITQSPPPYRVCVKHPYTGPSCSQVVPVCRPNPCQNGATCSRHKRSKFTACPD 180

QY 181 QFKGKFCBEGSDDCYVGDGYSGRGNRTVNHACLYWNSHLLLOENYMFMEDEATHGI 240

Db 181 QFKGKFCBEGSDDCYVGDGYSGRGNRTVNHACLYWNSHLLLOENYMFMEDEATHGI 240

QY 241 GEHNFCRNPDADKPCWCFIKVTNDKVWEYCDVSACSADQVAYPEESPTSTKLPGFDS 300

Db 241 GEHNFCRNPDADKPCWCFIKVTNDKVWEYCDVSACSADQVAYPEESPTSTKLPGFDS 300

QY 301 CGKTEIAERKIKRIYGGFKSTAGKHPWQASLOSSLPLTISMPQGHFCGGLIHPCWVLTA 360

Db 301 CGKTEIAERKIKRIYGGFKSTAGKHPWQASLOSSLPLTISMPQGHFCGGLIHPCWVLTA 360

QY 361 AHCTDIKTRHLKVVLGDQDLKKEEFHQSFVEKIFKYSHYNERDEIPHNDIALLLKPV 420

Db 361 AHCTDIKTRHLKVVLGDQDLKKEEFHQSFVEKIFKYSHYNERDEIPHNDIALLLKPV 420

QY 421 DGHCALESKYVKTVCCLPDGSPSGSECHISGMWVETGKGRQLLDKAKVLIANTLNSR 480

Db 421 DGHCALESKYVKTVCCLPDGSPSGSECHISGMWVETGKGRQLLDKAKVLIANTLNSR 480

QY 481 QLYDHMIDDSMICAGNLQKPGQDTCEKDGPTCEKDGTYVYVGVISWGLECGKRPVY 540

Db 481 QLYDHMIDDSMICAGNLQKPGQDTCEKDGPTCEKDGTYVYVGVISWGLECGKRPVY 540

QY 541 TQVTKFLNWKATIKSESSEGF 560

Db 541 TQVTKFLNWKATIKSESSEGF 560

RESULT 5

US-10-391-215-6

; Sequence 6, Application US/10391215

; Publication No. US20040009543A1

; GENERAL INFORMATION:

; APPLICANT: KIECHL, STEFAN

; APPLICANT: WILLEIT, JOHANN

; APPLICANT: WIEDERMANN, CHRISTIAN JOSEF

; APPLICANT: ROEMISCH, JUERGEN

; APPLICANT: WEIMER, THOMAS

; APPLICANT: FEUSSNER, ANNETTE

; APPLICANT: STOEHR, HANS-ARNOLD

; APPLICANT: DOERSAM, VOLKER

; TITLE OF INVENTION: MARBURG I MUTANT OF FACTOR VII ACTIVATING PROTEASE

; TITLE OF INVENTION: (FSAP) AS RISK FACTOR FOR ARTERIAL THROMBOSIS

; FILE REFERENCE: 06478.1457-01

; CURRENT FILING DATE: 2003-03-19

; CURRENT APPLICATION NUMBER: US/10/391.215

; PRIOR FILING DATE: 2001-07-26

; PRIOR FILING DATE: 2001-04-12

; NUMBER OF SEQ ID NOS: 8

; TITLE OF INVENTION: MUTANTS OF THE FACTOR VII-ACTIVATING PROTEASE AND
; FILE REFERENCE: 06478.1457
; CURRENT APPLICATION NUMBER: US/09/912.559
; CURRENT FILING DATE: 2001-07-26
; PRIOR APPLICATION NUMBER: DE 100 36 641.4
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: DE 100 50 040.4
; PRIOR FILING DATE: 2000-10-10
; PRIOR APPLICATION NUMBER: DE 100 52 319.6
; PRIOR FILING DATE: 2000-10-21
; PRIOR APPLICATION NUMBER: DE 101 18 706.8
; PRIOR FILING DATE: 2001-04-12
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 560
; TYPE: PRT
; ORGANISM: Homo sapiens
; ORGANISM: Homo sapiens
US-10-391-215-6

Query Match 99.7%; Score 3146; DB 15; Length 560;
Best Local Similarity 99.8%; Pred. No. 3.9e-240;
Matches 559; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MFARMSDLHVLNMAVGKTACGFSLSLLESIDPDWTPDQYDYSYEDYNOEENTTSITLT 60
DB 1 MFARMSDLHVLNMAVGKTACGFSLSLLESIDPDWTPDQYDYSYEDYNOEENTTSITLT 60
QY 61 HAENPDWYTTEDQADPCQNPCEHGGDCLVHGSTFTCSCLAPFSGNKCQKQVQNTCKDNPC 120
DB 61 HAENPDWYTTEDQADPCQNPCEHGGDCLVHGSTFTCSCLAPFSGNKCQKQVQNTCKDNPC 120
QY 121 GRGQCLITQSPPYRVCVKHPYTCPSQVVPVCRPNPCNGATCSRHKRSKFTACPD 180
DB 121 GRGQCLITQSPPYRVCVKHPYTCPSQVVPVCRPNPCNGATCSRHKRSKFTACPD 180
QY 181 QFKGKFCIGSDDCVVGDGYSYRGKQNRVTNQHACLYNWSHLLQENYNNMFMEDAEATHGI 240
DB 181 QFKGKFCIGSDDCVVGDGYSYRGKQNRVTNQHACLYNWSHLLQENYNNMFMEDAEATHGI 240
QY 241 GEHNFRCNPDADEKPCWCFIKVTNDKVWEYCDVSACSAQDVAYPEESPTSTKLPGFDS 300
DB 241 GEHNFRCNPDADEKPCWCFIKVTNDKVWEYCDVSACSAQDVAYPEESPTSTKLPGFDS 300
QY 301 CGKTEIAERKIKRIYGGFKSTAGKHPWQASLQSSLPITISMPQGHFCGGALIHPCWVILTA 360
DB 301 CGKTEIAERKIKRIYGGFKSTAGKHPWQASLQSSLPITISMPQGHFCGGALIHPCWVILTA 360
QY 361 AHCTDITKTHLKVVLGDQDLKKEEFHEQSFRVEKIFKYSYNERDEIPHNDIALLKLPV 420
DB 361 AHCTDITKTHLKVVLGDQDLKKEEFHEQSFRVEKIFKYSYNERDEIPHNDIALLKLPV 420
QY 421 DGHCALESKYVKTCLPDGSPFSGSECHISGNGVTETGKSRQLLDKAKVLIANTLCNSR 480
DB 421 DGHCALESKYVKTCLPDGSPFSGSECHISGNGVTETGKSRQLLDKAKVLIANTLCNSR 480
QY 481 QLYDHMIDDSMICAGNLQKPGQDTCQGDGSGPLTCEKDGTYVYVGIWSWGLECEKRPVY 540
DB 481 QLYDHMIDDSMICAGNLQKPGQDTCQGDGSGPLTCEKDGTYVYVGIWSWGLECEKRPVY 540
QY 541 TQVTKFLNWKATIKSES GF 560
DB 541 TQVTKFLNWKATIKSES GF 560

RESULT 6
US-09-912-559-4
; Sequence 4, Application US/09912559
; Patent No. US20020142316A1
; GENERAL INFORMATION:
; APPLICANT: ROEMISCH, JUERGEN
; APPLICANT: STOEHR, HANS-ARNOLD
; APPLICANT: FEUSSNER, ANNETTE
; APPLICANT: LANG, WIEGAND
; APPLICANT: WEIMER, THOMAS
; APPLICANT: BECKER, MARGRET
; APPLICANT: NERLICH, CLAUDIA
; APPLICANT: MUTH-NAUMANN, GUDRUN

; TITLE OF INVENTION: MUTANTS OF THE FACTOR VII-ACTIVATING PROTEASE AND
; FILE REFERENCE: 06478.1457
; CURRENT APPLICATION NUMBER: US/09/912.559
; CURRENT FILING DATE: 2001-07-26
; PRIOR APPLICATION NUMBER: DE 100 36 641.4
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: DE 100 50 040.4
; PRIOR FILING DATE: 2000-10-10
; PRIOR APPLICATION NUMBER: DE 100 52 319.6
; PRIOR FILING DATE: 2000-10-21
; PRIOR APPLICATION NUMBER: DE 101 18 706.8
; PRIOR FILING DATE: 2001-04-12
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 560
; TYPE: PRT
; ORGANISM: Homo sapiens
; ORGANISM: Homo sapiens
US-09-912-559-4

Query Match 99.7%; Score 3143; DB 9; Length 560;
Best Local Similarity 99.6%; Pred. No. 6.7e-240;
Matches 558; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MFARMSDLHVLNMAVGKTACGFSLSLLESIDPDWTPDQYDYSYEDYNOEENTTSITLT 60
DB 1 MFARMSDLHVLNMAVGKTACGFSLSLLESIDPDWTPDQYDYSYEDYNOEENTTSITLT 60
QY 61 HAENPDWYTTEDQADPCQNPCEHGGDCLVHGSTFTCSCLAPFSGNKCQKQVQNTCKDNPC 120
DB 61 HAENPDWYTTEDQADPCQNPCEHGGDCLVHGSTFTCSCLAPFSGNKCQKQVQNTCKDNPC 120
QY 121 GRGQCLITQSPPYRVCVKHPYTCPSQVVPVCRPNPCNGATCSRHKRSKFTACPD 180
DB 121 GRGQCLITQSPPYRVCVKHPYTCPSQVVPVCRPNPCNGATCSRHKRSKFTACPD 180
QY 181 QFKGKFCIGSDDCVVGDGYSYRGKQNRVTNQHACLYNWSHLLQENYNNMFMEDAEATHGI 240
DB 181 QFKGKFCIGSDDCVVGDGYSYRGKQNRVTNQHACLYNWSHLLQENYNNMFMEDAEATHGI 240
QY 241 GEHNFRCNPDADEKPCWCFIKVTNDKVWEYCDVSACSAQDVAYPEESPTSTKLPGFDS 300
DB 241 GEHNFRCNPDADEKPCWCFIKVTNDKVWEYCDVSACSAQDVAYPEESPTSTKLPGFDS 300
QY 301 CGKTEIAERKIKRIYGGFKSTAGKHPWQASLQSSLPITISMPQGHFCGGALIHPCWVILTA 360
DB 301 CGKTEIAERKIKRIYGGFKSTAGKHPWQASLQSSLPITISMPQGHFCGGALIHPCWVILTA 360
QY 361 AHCTDITKTHLKVVLGDQDLKKEEFHEQSFRVEKIFKYSYNERDEIPHNDIALLKLPV 420
DB 361 AHCTDITKTHLKVVLGDQDLKKEEFHEQSFRVEKIFKYSYNERDEIPHNDIALLKLPV 420
QY 421 DGHCALESKYVKTCLPDGSPFSGSECHISGNGVTETGKSRQLLDKAKVLIANTLCNSR 480
DB 421 DGHCALESKYVKTCLPDGSPFSGSECHISGNGVTETGKSRQLLDKAKVLIANTLCNSR 480
QY 481 QLYDHMIDDSMICAGNLQKPGQDTCQGDGSGPLTCEKDGTYVYVGIWSWGLECEKRPVY 540
DB 481 QLYDHMIDDSMICAGNLQKPGQDTCQGDGSGPLTCEKDGTYVYVGIWSWGLECEKRPVY 540
QY 541 TQVTKFLNWKATIKSES GF 560
DB 541 TQVTKFLNWKATIKSES GF 560

RESULT 7
US-10-391-215-8
; Sequence 8, Application US/10391215
; Publication No. US20040009543A1
; GENERAL INFORMATION:
; APPLICANT: KIECHL, STEFAN
; APPLICANT: WILLEIT, JOHANN

APPLICANT: WIEDERMANN, CHRISTIAN JOSEF
APPLICANT: ROEMISCH, JUERGEN
APPLICANT: WEIMER, THOMAS
APPLICANT: FEUGSNER, ANNETTE
APPLICANT: STOEHR, HANS-ARNOLD
APPLICANT: DOERSAM, VOLKER
TITLE OF INVENTION: MARBURG I MUTANT OF FACTOR VII ACTIVATING PROTEASE
FILE REFERENCE: 06478.1457-01
CURRENT APPLICATION NUMBER: US/10/391,215
CURRENT FILING DATE: 2003-03-19
PRIORITY APPLICATION NUMBER: 09/912,559
PRIORITY FILING DATE: 2001-07-26
PRIORITY FILING DATE: 2000-07-26
PRIORITY FILING DATE: 2000-10-10
PRIORITY FILING DATE: 2000-10-10
PRIORITY FILING DATE: 2000-10-21
PRIORITY FILING DATE: 2001-04-12
NUMBER OF SEQ ID NOS: 8
SOFTWARE: Patentin Ver. 2.1
LENGTH: 560
TYPE: PRT
ORGANISM: Homo sapiens
US-10-391-215-8

Query Match 99.7%; Score 3143; DB 15; Length 560;
Best Local Similarity 99.6%; Pred. No. 6.7e-240;
Matches 558; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MFARMSDLVLLMALVGTAGCFSLMSLESLLDPDQDYDYSDYEDYVQENSTSLT 60
DB 1 MFARMSDLVLLMALVGTAGCFSLMSLESLLDPDQDYDYSDYEDYVQENSTSLT 60

QY 61 HAENPDWYTEDQADPCQNPCEHGGDCLVHGSTFTCSCLAPFGSKNCKQVQNTCKDNPC 120
DB 61 HAENPDWYTEDQADPCQNPCEHGGDCLVHGSTFTCSCLAPFGSKNCKQVQNTCKDNPC 120

QY 121 GRGQCLITOSPYYRCVCKHPTGSCSQVPCRPNPQNGATCSRHKRSKFTACPD 180
DB 121 GRGQCLITOSPYYRCVCKHPTGSCSQVPCRPNPQNGATCSRHKRSKFTACPD 180

QY 181 QFKGKFCETGSDDCVYGDGYSYRGKMRVNOHACLYMNSHLLQENYNNMEDAETHGI 240
DB 181 QFKGKFCETGSDDCVYGDGYSYRGKMRVNOHACLYMNSHLLQENYNNMEDAETHGI 240

QY 241 GEHNFRCNPDADKPCWCFIKVNDKVKWEYCDVSACSAQDVAYPEESPTSTKLPGFDS 300
DB 241 GEHNFRCNPDADKPCWCFIKVNDKVKWEYCDVSACSAQDVAYPEESPTSTKLPGFDS 300

QY 301 CGKTEAERKIKRIYGGFKSTAGKHPQWASLSSPLTISMPQGHPCGALLHPCWVLT 360
DB 301 CGKTEAERKIKRIYGGFKSTAGKHPQWASLSSPLTISMPQGHPCGALLHPCWVLT 360

QY 361 AHCTDIKTRHLKVLGDQDLKKEEFHQSFVEKIFKYSHYNERDIPNDIALLKLPV 420
DB 361 AHCTDIKTRHLKVLGDQDLKKEEFHQSFVEKIFKYSHYNERDIPNDIALLKLPV 420

QY 421 DGHCALESKYVTKVCLPDGSPFSGSCHISGKVTETGKSGQLDQKVLANTLCNSR 480
DB 421 DGHCALESKYVTKVCLPDGSPFSGSCHISGKVTETGKSGQLDQKVLANTLCNSR 480

QY 481 QLYDHMIDDSMI CAGNLQKPGQDTCQDGGGPLTCEKDGTYYYGVISWGLECGKRPVY 540
DB 481 QLYDHMIDDSMI CAGNLQKPGQDTCQDGGGPLTCEKDGTYYYGVISWGLECGKRPVY 540

QY 541 TQVTKFLNWKATIKSESGF 560
DB 541 TQVTKFLNWKATIKSESGF 560

RESULT 8
US-10-172-712-28
Sequence 28, Application US/10172712
Publication No. US20030125232A1
GENERAL INFORMATION:
APPLICANT: GRIFFIN, JOHN H.
APPLICANT: GALE, ANDREW J.
APPLICANT: GETZOFF, ELIZABETH D.
APPLICANT: PELLEQUER, JEAN-LUC
TITLE OF INVENTION: STABILIZED PROTEINS WITH ENGINEERED DISULFIDE BONDS
FILE REFERENCE: 4198-4001US1
CURRENT APPLICATION NUMBER: US/10/172,712
CURRENT FILING DATE: 2002-09-30
PRIORITY APPLICATION NUMBER: 60/298,578
PRIORITY FILING DATE: 2001-06-14
NUMBER OF SEQ ID NOS: 32
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 28
LENGTH: 655
TYPE: PRT
ORGANISM: Homo sapiens
US-10-172-712-28

Query Match 26.4%; Score 832.5; DB 14; Length 655;
Best Local Similarity 33.2%; Pred. No. 3.9e-57;
Matches 192; Conservative 87; Mismatches 216; Indels 83; Gaps 21;

QY 31 ESLLDPDWTDPDYDYSDYEDYVQENSTSLTHAENPDW-----YVTEQDA----- 74
DB 99 QALTEGRPCRFPRYGRGRLHACTSEGSAHRK---WCATTHYDRDRAWGYCVAETPPP 155

QY 75 -----DPCQNPCEHGGDCLVHGSTFTCSCLAPFGSKNCKQVQNTCKDNPCGRGQCL 126
DB 156 GGAALDPCASGCLNGGSCNTQDPQSYHSCSPRAFTG-----KD--CGTEKCF 203

QY 127 ITGSPPIY-----RCVCKHPTGSCSQV--PVCPRNPQNGATCSRHK 169
DB 204 DETRYELEGDRWARVRQGHVEQCEC---FGRTWCEGTRHTACLSPPCLNGGTCHLIV 260

QY 170 RRSKFTCACPDQKFKGFCETGSDDCVYGDGYSYRGKMRVNOHACLYMNSHLLQENY 228
DB 261 ATGTTVCACPPGAGRLCNIEPDERCFLNGTCYRGVASTSASGLSCLANNSDLLYQELH 320

QY 229 NMFMEDAEETHGIGBHNFCRNPDADEKPCWCFIKVNDKVKWEYCDVSACSA--QDVAYPEE 286
DB 321 VDSVGAALLGLGPHAYCRNPNDNERPCVY--VKDSALSWEYCRLEACESLTRVQLSPDL 379

QY 287 SPTSPSTKLPQFSCGKTEAERKIK--RIYGGFKSTAGKHPQWASLSSPLTISMPQGH 345
DB 380 LATLPEPASPCQRQACRRHKRTFLRPRIIGSSSLPGSHPLAAIYG-----DS 430

QY 346 FCGGALLHPCWVLTAAHCTDIKTRH--LKVLGDQDLKKEEFHQSFVEKIFKYSHYNE 403
DB 431 FCAGSLVHTCWVYSAACHFSGHSPRDSVSVLQGHFFNRTDTVTQFGIEKIPYTLYSV 490

QY 404 RDETPHNDIALLKLPVDGHCALSKYVTKVCLPD--GSPPSGSECHISGK--VTEGKG 460
DB 491 FNPSDH--DLVILRLKKGDRCATRSQVQVFCILPEGSGTFPAGHKQCIAGHLDENVSG 549

QY 461 -SRQLLDKAKVLANTLCNSRQLYDHMIDDSMI CAGNLQKPGQDTCQDGGGPLTCEKDG 519
DB 550 YSSSLREALVPLVADHCKSSPEVYGADISPNMLCAGYFDCK--SDACQDGGGGLACEKNG 608

QY 520 TYYVYGVISWGLECGK--RFGVTVTKFLNWKATIK 555
DB 609 VAYLYGIISGDDGCGRLHFKGVTVRVANYVDWINDIR 646

RESULT 9
US-09-974-298-145
Sequence 145, Application US/09974298
Patent No. US20020156263A1

GENERAL INFORMATION:
; APPLICANT: Chen, Huei-Mei
; TITLE OF INVENTION: GENES EXPRESSED IN BREAST CANCER
; FILE REFERENCE: PA-0037 P
; CURRENT APPLICATION NUMBER: US/09/974,299
; CURRENT FILING DATE: 2001-10-04
; PRIOR APPLICATION NUMBER: 60/238,331
; PRIOR FILING DATE: 2000-05-10
; NUMBER OF SEQ ID NOS: 194
; SOFTWARE: PERL Program
; SEQ ID NO 145
; LENGTH: 562
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20020156263A1 1001470CD1
US-09-974-298-145

Query Match 24.3%; Score 767.5; DB 9; Length 562;
Best Local Similarity 35.0%; Pred. No. 4.4e-52;
Matches 182; Conservative 77; Mismatches 176; Indels 85; Gaps 20;

QY 77 CQNPCEHGGDC--LVHGSTFTCSCLAPFSGNKCO-KVQNTCKDNP--CGRG-----123
Db 86 CSEPRCFNGGTCQOALYFSDVFCQCPGFAKCCCEIDTRATCYEDQGISYRGTWSTAESG 145
QY 124 -QCLITQSPPYRCVKHPYTGPCSCQVVPVCRPNPCQ-----NGATCSRHRKRSKFTCAC 178
Db 146 AECTWNNS-----SALAQKPYSGR-----RPDAIRLGLGNHNYCRNPDNRDPSKPCYV 192
QY 179 PDQFK-GKF-----CEIGSDDCYVGDGYSYRGKMNRTVNOHACLYWNSHLLQENY 228
Db 193 ---FRAGKYSSEFCSTPACSEGNDCYFGNSAYRGTHSLTESGASCLRNWSMILIGKVY 249
QY 229 NMFMEDATHGIGHNFCRNPDADEKPCWCFIKVTNDKWKWEYCDVSACSADQVAYPEESP 288
Db 250 TAQNPQAQALGLGKHNYCRNPDGAKPWCHV-LXNRRLTWECYCDVPSCS-----297
QY 289 TEPSTKLPGFDSGCKTEIAERKIKRIYGGFKSTAGKHPQASLOSLLPLTISMPQGHFCG 348
Db 298 -----TCGLRQYSQPF-RIKGGLFADIASHPWQAIAFAK--HRRSPGERFLCG 343
QY 349 GALIHPCWVLTAAHCTD--IKTRHLKVLGQDLKKEEFHQSPRVEKIFKYSHYNERDE 406
Db 344 GILISSCWILSAACFCQERFPFPHLTVLIGTRYVWPGEEBQKPEVEKIYVHKFED--DD 401
QY 407 IPHNDIALLLKVPDGHCALESKYVKTCLP--DGSPPSGSECHISGMVTEITCKG--SR 462
Db 402 TYDNDIALLLQKSDSSRCAQESSVVRTVCLPPADLPDWTCELSGYGKEALSPFYSE 461
QY 463 QLLDAKVLIANTLNSRQLYDHMIDDSMI CAGNLQKPG-----QDTCQSDSGGFLTCEK 517
Db 462 RLKEAHVRLYPSRRTCSQHLNRTVTDNMLCAGDTRSGGPOANLHDACQSDSGGFLVCLN 521
QY 518 DGTYYVYGVSVGLECGKR--PGVYTVTKFLNMIKATIK 555
Db 522 DGRMTLVGIISGLGCGQKDPGVYTKVTNYLDWIRDNR 561

RESULT 10
US-09-987-457-18
; Sequence 18, Application US/09987457
; Publication No. US20030013150A1
; GENERAL INFORMATION:
; APPLICANT: Manosroi, Aranya
; APPLICANT: Manosroi, Jiradej
; APPLICANT: Tayapiwatana, Chatchai
; APPLICANT: Goetz, Friedrich
; APPLICANT: Werner, Rolf-Guenther
; TITLE OF INVENTION: Methods for Large Scale Protein Production in Prokaryotes
; FILE REFERENCE: 0652.2180001
; CURRENT APPLICATION NUMBER: US/09/987,457

GENERAL INFORMATION:
; APPLICANT: Chen, Huei-Mei
; TITLE OF INVENTION: GENES EXPRESSED IN BREAST CANCER
; FILE REFERENCE: PA-0037 P
; CURRENT APPLICATION NUMBER: US/09/974,299
; CURRENT FILING DATE: 2001-10-04
; PRIOR APPLICATION NUMBER: 60/238,331
; PRIOR FILING DATE: 2000-05-10
; NUMBER OF SEQ ID NOS: 194
; SOFTWARE: PERL Program
; SEQ ID NO 145
; LENGTH: 562
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20020156263A1 1001470CD1
US-09-974-298-145

Query Match 24.3%; Score 766.5; DB 10; Length 527;
Best Local Similarity 35.0%; Pred. No. 4.9e-52;
Matches 182; Conservative 77; Mismatches 176; Indels 85; Gaps 20;

QY 77 CQNPCEHGGDC--LVHGSTFTCSCLAPFSGNKCO-KVQNTCKDNP--CGRG-----123
Db 51 CSEPRCFNGGTCQOALYFSDVFCQCPGFAKCCCEIDTRATCYEDQGISYRGTWSTAESG 110
QY 124 -QCLITQSPPYRCVKHPYTGPCSCQVVPVCRPNPCQ-----NGATCSRHRKRSKFTCAC 178
Db 111 AECTWNNS-----SALAQKPYSGR-----RPDAIRLGLGNHNYCRNPDNRDPSKPCYV 157
QY 179 PDQFK-GKF-----CEIGSDDCYVGDGYSYRGKMNRTVNOHACLYWNSHLLQENY 228
Db 158 ---FRAGKYSSEFCSTPACSEGNDCYFGNSAYRGTHSLTESGASCLRNWSMILIGKVY 214
QY 229 NMFMEDATHGIGHNFCRNPDADEKPCWCFIKVTNDKWKWEYCDVSACSADQVAYPEESP 288
Db 215 TAQNPQAQALGLGKHNYCRNPDGAKPWCHV-LXNRRLTWECYCDVPSCS-----262
QY 289 TEPSTKLPGFDSGCKTEIAERKIKRIYGGFKSTAGKHPQASLOSLLPLTISMPQGHFCG 348
Db 263 -----TCGLRQYSQPF-RIKGGLFADIASHPWQAIAFAK--HRRSPGERFLCG 308
QY 349 GALIHPCWVLTAAHCTD--IKTRHLKVLGQDLKKEEFHQSPRVEKIFKYSHYNERDE 406
Db 309 GILISSCWILSAACFCQERFPFPHLTVLIGTRYVWPGEEBQKPEVEKIYVHKFED--DD 366
QY 407 IPHNDIALLLKVPDGHCALESKYVKTCLP--DGSPPSGSECHISGMVTEITCKG--SR 462
Db 367 TYDNDIALLLQKSDSSRCAQESSVVRTVCLPPADLPDWTCELSGYGKEALSPFYSE 426
QY 463 QLLDAKVLIANTLNSRQLYDHMIDDSMI CAGNLQKPG-----QDTCQSDSGGFLTCEK 517
Db 427 RLKEAHVRLYPSRRTCSQHLNRTVTDNMLCAGDTRSGGPOANLHDACQSDSGGFLVCLN 486
QY 518 DGTYYVYGVSVGLECGKR--PGVYTVTKFLNMIKATIK 555
Db 487 DGRMTLVGIISGLGCGQKDPGVYTKVTNYLDWIRDNR 526

RESULT 11
US-09-987-455-19
; Sequence 19, Application US/09987455
; Publication No. US20030049729A1
; GENERAL INFORMATION:
; APPLICANT: Aranya Manosroi
; APPLICANT: Jiradej Manosroi
; APPLICANT: Chatchai Tayapiwatana
; APPLICANT: Friedrich Goetz
; APPLICANT: Rolf-Guenther Werner
; TITLE OF INVENTION: Methods for Large Scale Production of Recombinant
; FILE REFERENCE: 0652.2190001
; CURRENT APPLICATION NUMBER: US/09/987,455
; CURRENT FILING DATE: 2001-11-14
; PRIOR APPLICATION NUMBER: 60/268,574
; PRIOR FILING DATE: 2001-02-15
; PRIOR APPLICATION NUMBER: GB 0027779.8
; PRIOR FILING DATE: 2000-11-14

; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 19
; LENGTH: 527
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-987-455-19

Query Match 24.3%; Score 766.5; DB 10; Length 527;
Best Local Similarity 35.0%; Pred. No. 4.9e-52;
Matches 182; Conservative 77; Mismatches 176; Indels 85; Gaps 20;

QY 77 CQNPCEHGGDC--LVHGSTFTCSCLAPFSGNKKQ-KVQNTCKDNP--CGRG-----123
DB 51 CSEPRCFNGGTCQOALYFSDFCQCEIDTRATCYEDQGISYRGTWSTAESG 110
QY 124 -QCLITQSPYRCVKHPYTGPSQSVVPCRNPCQ-----NGATCSRHKRSKFTCAC 178
DB 111 AECTWNWS-----SALAQKPYSGR-----RPAIRLGLGNHNYCRNPDSDSKPCYV 157
QY 179 PDQFK-GKF-----CEIGSDDCYVGDGYSYRGKMNRTVNOHACLYWNSHLLQENY 228
DB 158 ---FKAGKYSSEFCSTPACSEGNSDCYFGNSAYRGTHSLTESGASCLPWNMILIGKYV 214
QY 229 NMVEDAETHGIEHNFNCPNPDADKPMCFIKVTNDKVKWEYCDVSACSADQVAYPEESP 288
DB 215 TAQNPQAQALGLGKHNKCRNPDGDAKPWCHV-LKNRLTWECYDVPSCS-----262
QY 289 TEPSTKLPGDFSCGKTEIAERKIKRIYGGFKSTAGKHPWQASLQSSLPITISMPQGHFCG 348
DB 263 -----TCGLRQYSQPOF-RIKGLFADIASHPWQAIFAK--HRRSPGERFLCG 308
QY 349 GALIHPCWVLTAAHCTD--IKTRHLKVLGDQDLKKEEFHQSFVEKIFKYSHYNERDE 406
DB 309 GILISSCWILSAAHCFQERFPHPHLTVILGRYRVVPGEEQKFEVEKYIVHKEFD--DD 366
QY 407 IPHNDIALKLPVDGHCALSKYKVTCLP--DGSFPGSECHISGWTGTGKG--SR 462
DB 367 TYNDIALLOKSDSRCAQESSVVRTCLPADLQLPDWTCELSGYGKHEALSPFYSE 426
QY 463 QLLDAKVKLIANTLNSRQLYDHMIDDSMICAGNLQKPG-----QDTQCGSDGGPLTCEK 517
DB 427 RLKEAHRVLYPSSRCTSQHLLNRTVTDNMLCAGDTRSGGPQANLHDAQCQSDGGPLVCLN 486
QY 518 DGTYYVYGVISWGLECGKR--PGVYTVTKFLNWKATIK 555
DB 487 DGRMTLVGIISWGLGCGKQDVPGVYTKVNYLDWIRDNR 526

RESULT 13
US-10-360-101-203
; Sequence 203, Application US/10360101
; Publication No. US20040009550A1
; GENERAL INFORMATION:
; APPLICANT: Moll, Gert N.
; TITLE OF INVENTION: Export and modification of (poly)peptide in the lantibiotic way
; FILE REFERENCE: 2183-5673
; CURRENT APPLICATION NUMBER: US/10/360,101
; CURRENT FILING DATE: 2003-02-07
; PRIOR APPLICATION NUMBER: EP 02077060.8
; PRIOR FILING DATE: 2002-05-24
; NUMBER OF SEQ ID NOS: 309
; SOFTWARE: Patent In version 3.1
; SEQ ID NO 203
; LENGTH: 527
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: sequence of alteplase
US-10-360-101-203

Query Match 24.3%; Score 766.5; DB 15; Length 527;
Best Local Similarity 35.0%; Pred. No. 4.9e-52;
Matches 182; Conservative 77; Mismatches 176; Indels 85; Gaps 20;

QY 77 CQNPCEHGGDC--LVHGSTFTCSCLAPFSGNKKQ-KVQNTCKDNP--CGRG-----123
DB 51 CSEPRCFNGGTCQOALYFSDFCQCEIDTRATCYEDQGISYRGTWSTAESG 110

; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 19
; LENGTH: 527
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-987-455-19

Query Match 24.3%; Score 766.5; DB 10; Length 527;
Best Local Similarity 35.0%; Pred. No. 4.9e-52;
Matches 182; Conservative 77; Mismatches 176; Indels 85; Gaps 20;

QY 77 CQNPCEHGGDC--LVHGSTFTCSCLAPFSGNKKQ-KVQNTCKDNP--CGRG-----123
DB 51 CSEPRCFNGGTCQOALYFSDFCQCEIDTRATCYEDQGISYRGTWSTAESG 110
QY 124 -QCLITQSPYRCVKHPYTGPSQSVVPCRNPCQ-----NGATCSRHKRSKFTCAC 178
DB 111 AECTWNWS-----SALAQKPYSGR-----RPAIRLGLGNHNYCRNPDSDSKPCYV 157
QY 179 PDQFK-GKF-----CEIGSDDCYVGDGYSYRGKMNRTVNOHACLYWNSHLLQENY 228
DB 158 ---FKAGKYSSEFCSTPACSEGNSDCYFGNSAYRGTHSLTESGASCLPWNMILIGKYV 214
QY 229 NMVEDAETHGIEHNFNCPNPDADKPMCFIKVTNDKVKWEYCDVSACSADQVAYPEESP 288
DB 215 TAQNPQAQALGLGKHNKCRNPDGDAKPWCHV-LKNRLTWECYDVPSCS-----262
QY 289 TEPSTKLPGDFSCGKTEIAERKIKRIYGGFKSTAGKHPWQASLQSSLPITISMPQGHFCG 348
DB 263 -----TCGLRQYSQPOF-RIKGLFADIASHPWQAIFAK--HRRSPGERFLCG 308
QY 349 GALIHPCWVLTAAHCTD--IKTRHLKVLGDQDLKKEEFHQSFVEKIFKYSHYNERDE 406
DB 309 GILISSCWILSAAHCFQERFPHPHLTVILGRYRVVPGEEQKFEVEKYIVHKEFD--DD 366
QY 407 IPHNDIALKLPVDGHCALSKYKVTCLP--DGSFPGSECHISGWTGTGKG--SR 462
DB 367 TYNDIALLOKSDSRCAQESSVVRTCLPADLQLPDWTCELSGYGKHEALSPFYSE 426
QY 463 QLLDAKVKLIANTLNSRQLYDHMIDDSMICAGNLQKPG-----QDTQCGSDGGPLTCEK 517
DB 427 RLKEAHRVLYPSSRCTSQHLLNRTVTDNMLCAGDTRSGGPQANLHDAQCQSDGGPLVCLN 486
QY 518 DGTYYVYGVISWGLECGKR--PGVYTVTKFLNWKATIK 555
DB 487 DGRMTLVGIISWGLGCGKQDVPGVYTKVNYLDWIRDNR 526

RESULT 12
US-10-432-842-1
; Sequence 1, Application US/10432842
; Publication No. US20040071707A1
; GENERAL INFORMATION:
; APPLICANT: Veronica A. CARROLL
; APPLICANT: Adrian L. HARRIS
; APPLICANT: Roy BICKNELL
; APPLICANT: Pat PRICE
; TITLE OF INVENTION: MODULATION OF CELL GROWTH
; FILE REFERENCE: 117-450 / N.79507A SER
; CURRENT APPLICATION NUMBER: US/10/432,842
; CURRENT FILING DATE: 2003-09-27
; PRIOR APPLICATION NUMBER: PCT/GB01/05244
; PRIOR FILING DATE: 2001-11-28
; PRIOR APPLICATION NUMBER: GB 0029001.5
; PRIOR FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: MS Word
; SEQ ID NO 1
; LENGTH: 527
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-432-842-1

QY 124 -QCLITQSPYRCVCKHPYTGSCSQVVPVCRPNFCQ-----NGATCSHRRSKFTAC 178
DB 111 AECTNWS-----SALAKPYSGR-----RPDAIRLGLGNHNYCRNPRDRSKFCV 157
QY 179 PDQFK-GKF-----CEIGDDCYVDGYSYRGKMRVTHQACLYWNSHLLLOENY 228
DB 158 ---FKAGKYSSEFCSTPACSEGNDCYFGNGSAYRGTHSLTESGASCLPWNMILIGKY 214
QY 229 NMFMEDAETHGIEHNPCRNPDADKPWCFIKVTNDKWKVEYCDVSACSADQVAYPEESP 288
DB 215 TAQNPSAALGLGKHNVCNPDGDAKPWCHV-LKNRLTWECYDVFSCS-----262
QY 289 TEPSTKLPFGDSCGKTEIAERKIKRIYGGFKSTAGKHPWQASLOSLLPLTISMPQGHFCG 348
DB 263 -----TCGLRQYSQPF-RIKGGLFADIASHPWQAIFAK--HRRSPGERFLCG 308
QY 349 GALHPCWVLTAAHCTD--IKTRHLKVVLGDQDLKKEEFHQSFVEKIFKYSHYNERDE 406
DB 309 GILISSCWILSAACFOERPPPHLTIVLGRVTVVPGEEQKFEVEKIVHKEFD--DD 366
QY 407 IPHNDIALKLKPVVDGHCALSKYKVTCLP--DGSFPGSECHISGWGVTETGKG--SR 462
DB 367 TYNDIALQLKSDSSRCAQESSVVRTVCLPPADLQLPDWTECELSGYGKHEALSPFYSE 426
QY 463 QLLDAKVKLIANTLCNSRQLYDHMDSDMICAGNLQKPG-----QDTCQDSDSGPLTCEK 517
DB 427 RLKEAHVRLYPSRCTQHLNLRVTDNMLCAGDTRSGGPGQANLHDAACQDSDSGPLVCLN 486
QY 518 DGTYYVYGVISWGLECGKR--PGVYTVQTKPLNMIKATIK 555
DB 487 DGRMTLVGIISWGLCGQKQVPGVYTKVTNYLDWIRDNR 526

RESULT 14
US-09-969-271-7
; Sequence 7, Application US/09969271
; Patent No. US20020098179A1
; GENERAL INFORMATION:
; APPLICANT: Pfizer Inc. (All designated States except GB and EP (GB));
; APPLICANT: Pfizer Limited (GB and EP (GB) only)
; TITLE OF INVENTION: Pharmaceutical Combinations
; FILE REFERENCE: PCS10951APME
; CURRENT APPLICATION NUMBER: US/09/969,271
; PRIOR FILING DATE: 2001-10-01
; PRIOR APPLICATION NUMBER: GB 0025473.0
; PRIOR FILING DATE: 2000-10-17
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 562
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-969-271-7

Query Match 24.3%; Score 766.5; DB 9; Length 562;
Best Local Similarity 35.0%; Pred. No. 5.3e-52;
Matches 182; Conservative 77; Mismatches 176; Indels 85; Gaps 20;
QY 77 CQNPCEHGGDC--LVHGSTFTCSCLAPFSGNKKQ-KVQNTCKDNP--CGRG-----123
DB 86 CSEPRCFNGGTCQOALYFSDFCVQCPEGFAGKCCCEIDTRATCYEDQGISYRGWTSTAESG 145
QY 124 -QCLITQSPYRCVCKHPYTGSCSQVVPVCRPNFCQ-----NGATCSHRRSKFTAC 178
DB 146 AECTNWS-----SALAKPYSGR-----RPDAIRLGLGNHNYCRNPRDRSKFCV 192
QY 179 PDQFK-GKF-----CEIGDDCYVDGYSYRGKMRVTHQACLYWNSHLLLOENY 228
DB 193 ---FKAGKYSSEFCSTPACSEGNDCYFGNGSAYRGTHSLTESGASCLPWNMILIGKY 249
QY 229 NMFMEDAETHGIEHNPCRNPDADKPWCFIKVTNDKWKVEYCDVSACSADQVAYPEESP 288
DB 250 TAQNPSAALGLGKHNVCNPDGDAKPWCHV-LKNRLTWECYDVFSCS-----297

QY 289 TEPSTKLPFGDSCGKTEIAERKIKRIYGGFKSTAGKHPWQASLOSLLPLTISMPQGHFCG 348
DB 298 -----TCGLRQYSQPF-RIKGGLFADIASHPWQAIFAK--HRRSPGERFLCG 343
QY 349 GALHPCWVLTAAHCTD--IKTRHLKVVLGDQDLKKEEFHQSFVEKIFKYSHYNERDE 406
DB 344 GILISSCWILSAACFOERPPPHLTIVLGRVTVVPGEEQKFEVEKIVHKEFD--DD 401
QY 407 IPHNDIALKLKPVVDGHCALSKYKVTCLP--DGSFPGSECHISGWGVTETGKG--SR 462
DB 402 TYNDIALQLKSDSSRCAQESSVVRTVCLPPADLQLPDWTECELSGYGKHEALSPFYSE 461
QY 463 QLLDAKVKLIANTLCNSRQLYDHMDSDMICAGNLQKPG-----QDTCQDSDSGPLTCEK 517
DB 462 RLKEAHVRLYPSRCTQHLNLRVTDNMLCAGDTRSGGPGQANLHDAACQDSDSGPLVCLN 521
QY 518 DGTYYVYGVISWGLECGKR--PGVYTVQTKPLNMIKATIK 555
DB 522 DGRMTLVGIISWGLCGQKQVPGVYTKVTNYLDWIRDNR 561

RESULT 15
US-10-411-037-26
; Sequence 26, Application US/10411037
; Publication No. US20040043446A1
; GENERAL INFORMATION:
; APPLICANT: Neose Technologies, Inc.
; APPLICANT: Defrees, Shawn
; APPLICANT: Zopf, David
; APPLICANT: Bayer, Robert
; APPLICANT: Hakes, David
; APPLICANT: Chen, Xi
; APPLICANT: Bowe, Caryn
; TITLE OF INVENTION: ALPHA GALACTOSIDASE A: REMODELING AND GLYCOCONJUGATION OF ALPHA
; TITLE OF INVENTION: GALACTOSIDASE A
; FILE REFERENCE: 040853-01-5082
; CURRENT APPLICATION NUMBER: US/10/411,037
; CURRENT FILING DATE: 2003-04-09
; PRIOR APPLICATION NUMBER: US 60/328,523
; PRIOR FILING DATE: 2001-10-10
; PRIOR APPLICATION NUMBER: US 60/344,692
; PRIOR FILING DATE: 2001-10-19
; PRIOR APPLICATION NUMBER: US 60/387,292
; PRIOR FILING DATE: 2002-08-07
; PRIOR APPLICATION NUMBER: US 60/391,777
; PRIOR FILING DATE: 2002-06-25
; PRIOR APPLICATION NUMBER: US 60/396,594
; PRIOR FILING DATE: 2002-07-17
; PRIOR APPLICATION NUMBER: US 60/404,249
; PRIOR FILING DATE: 2002-08-16
; PRIOR APPLICATION NUMBER: US 60/407,527
; PRIOR FILING DATE: 2002-08-28
; NUMBER OF SEQ ID NOS: 75
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 26
; LENGTH: 562
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-411-037-26

Query Match 24.3%; Score 766.5; DB 12; Length 562;
Best Local Similarity 35.0%; Pred. No. 5.3e-52;
Matches 182; Conservative 77; Mismatches 176; Indels 85; Gaps 20;
QY 77 CQNPCEHGGDC--LVHGSTFTCSCLAPFSGNKKQ-KVQNTCKDNP--CGRG-----123
DB 86 CSEPRCFNGGTCQOALYFSDFCVQCPEGFAGKCCCEIDTRATCYEDQGISYRGWTSTAESG 145
QY 124 -QCLITQSPYRCVCKHPYTGSCSQVVPVCRPNFCQ-----NGATCSHRRSKFTAC 178
DB 146 AECTNWS-----SALAKPYSGR-----RPDAIRLGLGNHNYCRNPRDRSKFCV 192

[illegible]

Search completed: May 24, 2004, 09:58:57
Job time : 44 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 24, 2004, 09:39:54 ; Search time 54 Seconds
(without alignments)
2930.122 Million cell updates/sec

Title: US-09-912-559-4
Perfect score: 3153
Sequence: 1 MFARMSDLVLLMALVGKT.....TQVTKFLNLIKATIKESG 560

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:*

- 1: Geneseqp1980s:*
- 2: Geneseqp1990s:*
- 3: Geneseqp2000s:*
- 4: Geneseqp2001s:*
- 5: Geneseqp2002s:*
- 6: Geneseqp2003as:*
- 7: Geneseqp2003bs:*
- 8: Geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	3153	100.0	560	AA017145	AA017145 Human blo
2	3143	99.7	560	AA042484	AA042484 Human ORF
3	3143	99.7	560	AA017144	AA017144 Human blo
4	822.5	26.1	655	AA053962	AA053962 Hepatocyt
5	821.5	26.1	655	AA089197	AA089197 Human hep
6	801.5	25.4	441	AA070643	AA070643 Modified
7	801.5	25.4	473	AA070642	AA070642 Modified
8	796.5	25.3	689	AA070850	AA070850 Novel pro
9	791	25.1	439	AA09360	AA09360 t-PA vari
10	787.5	25.0	434	AA013922	AA013922 Delta (92
11	786	24.9	439	AA09261	AA09261 t-PA vari
12	784	24.9	439	AA09258	AA09258 t-PA vari
13	783	24.8	439	AA09259	AA09259 t-PA vari
14	782	24.8	623	AA017142	AA017142 Ile(277)t
15	781	24.8	623	AA017140	AA017140 t-PA with
16	775	24.6	527	AA070858	AA070858 Human t-P
17	775	24.6	562	AA041545	AA041545 Modified
18	774	24.5	518	AA074684	AA074684 t-PA mute
19	774	24.5	522	AA013917	AA013917 Delta (46
20	773	24.5	527	AA070893	AA070893 Human t-P
21	772	24.5	521	AA074678	AA074678 t-PA mute
22	771.5	24.5	483	AA09257	AA09257 t-PA vari
23	771	24.5	522	AA074683	AA074683 t-PA vari
24	770.5	24.4	483	AA09252	AA09252 t-PA vari
25	770	24.4	519	AA074687	AA074687 t-PA vari

26	770	24.4	527	2	AA070861	AA070861 Human t-P
27	770	24.4	527	2	AA070856	AA070856 Human t-P
28	770	24.4	527	2	AA070869	AA070869 Human t-P
29	769.5	24.4	483	2	AA09246	AA09246 t-PA vari
30	769	24.4	519	2	AA074680	AA074680 t-PA mute
31	769	24.4	522	2	AA013920	AA013920 Delta (46
32	769	24.4	527	2	AA070874	AA070874 Human t-P
33	769	24.4	527	2	AA070859	AA070859 Human t-P
34	768.5	24.4	483	2	AA070889	AA070889 Human t-P
35	768	24.4	522	2	AA074685	AA074685 t-PA mute
36	768	24.4	527	2	AA070896	AA070896 Human t-P
37	768	24.4	527	2	AA070904	AA070904 Human t-P
38	768	24.4	527	2	AA070891	AA070891 Human t-P
39	767.5	24.3	483	2	AA09250	AA09250 t-PA vari
40	767.5	24.3	483	2	AA09251	AA09251 t-PA vari
41	767.5	24.3	562	2	AA023802	AA023802 Zymogen-1
42	767	24.3	518	2	AA074679	AA074679 t-PA mute
43	767	24.3	522	2	AA014486	AA014486 Delta (46
44	767	24.3	527	2	AA070876	AA070876 Human t-P
45	767	24.3	527	2	AA070865	AA070865 Human t-P

ALIGNMENTS

RESULT 1
AA017145
ID AAO17145 standard; protein; 560 AA.

XX	AA017145;
XX	13-JUN-2002 (first entry)
XX	Human blood coagulation factor VII activating protease mutant.
XX	Human; blood coagulation factor VII activating protease; FSAP;
KW	single-chain plasminogen activator; bleeding disorder; haematological;
KW	haemostatic; mutant; mutin.
OS	Homo sapiens.
OS	Synthetic.
XX	EPI182258-A1.
XX	27-FEB-2002.
XX	05-JUL-2001; 2001EP-00115691.
XX	26-JUL-2000; 2000DE-01036641.
PR	10-OCT-2000; 2000DE-01050040.
PR	21-OCT-2000; 2000DE-01052319.
PR	12-APR-2001; 2001DE-01018706.
XX	(AVET) AVENTIS BEHRING GMBH.
XX	Roemisch J, Stoehr H, Feussner A, Lang W, Weimer T, Becker M;
PI	Nerlich C, Muth-Naumann G;
XX	WPI; 2002-270939/32.
DR	N-PSDE; AAL45697.
XX	New nucleic acid encoding mutant factor 7 activating protease, useful for
PT	diagnosis, treatment and prevention of coagulation disorders, also
PT	related protein and antibodies.
XX	Disclosure; Page 20-22; 27pp; German.
PS	The present invention relates to a mutant of the DNA sequence encoding
XX	the protease (FSAP) that activates blood coagulation factor VII (FVII)
CC	and single-chain plasminogen activator, where at least one of the base
CC	changes G to C at nucleotide 1177 and G to A at nucleotide 1601 is
CC	present. The mutant sequences can be used in the treatment and prevention
CC	of bleeding disorders associated with inherited or acquired defects of

CC blood coagulation factors V, VIII, IX, X, XI, XII, von Willebrand's
 CC factor and/or with antibodies against any of these proteins. The present
 CC sequence is the mutant human FSAP protein
 XX
 SQ Sequence 560 AA;

Query Match 100.0%; Score 3153; DB 5; Length 560;
 Best Local Similarity 100.0%; Pred. No. 7.5e-186; Indels 0; Gaps 0;
 Matches 560; Conservative 0; Mismatches 0;

QY 1 MFARMSDLHVLMLVKGKTACGFSLSLSLSDPDWTPDQYDYSYEDYNOEENTSSLT 60
 DB 1 MFARMSDLHVLMLVKGKTACGFSLSLSLSDPDWTPDQYDYSYEDYNOEENTSSLT 60
 QY 61 HAENPDWYTEDQADPCQPNPCBHGDCVHSGTFTCSCLAPFSGNKCQKQVNTCKDNPC 120
 DB 61 HAENPDWYTEDQADPCQPNPCBHGDCVHSGTFTCSCLAPFSGNKCQKQVNTCKDNPC 120
 QY 121 GRGCLITQSPFYRCVCKHPYTGPCSCQVVPVCRPNPCQNGATCSRHKRSKFTACPD 180
 DB 121 GRGCLITQSPFYRCVCKHPYTGPCSCQVVPVCRPNPCQNGATCSRHKRSKFTACPD 180
 QY 181 QFKGKFCIEIGSDCYVGDGYSYRGKQNRVTVNOHACLYNWSHLLLOENYMFMEDEATHGI 240
 DB 181 QFKGKFCIEIGSDCYVGDGYSYRGKQNRVTVNOHACLYNWSHLLLOENYMFMEDEATHGI 240
 QY 241 GEHNFRCNPDADEKPCWCFIKVTNDKVKWEYCDVSACSADQVAYPESPTSTKLPGPS 300
 DB 241 GEHNFRCNPDADEKPCWCFIKVTNDKVKWEYCDVSACSADQVAYPESPTSTKLPGPS 300
 QY 301 CGTEIAERIKRIYGGFKSTAGKHPQASLOSSLPITISMPQGHFCGGLIHPQWVLT 360
 DB 301 CGTEIAERIKRIYGGFKSTAGKHPQASLOSSLPITISMPQGHFCGGLIHPQWVLT 360
 QY 361 AHCTDIKTRHLKVLGDQDLKKEEFHQSFQVQKIPKYSHYNERDEIPHNDIALKLPV 420
 DB 361 AHCTDIKTRHLKVLGDQDLKKEEFHQSFQVQKIPKYSHYNERDEIPHNDIALKLPV 420
 QY 421 DGHCALESKVKVTCLPDGSPFSGSECHI:SGMGVTEGTGSRQLDQKLIANTLNSR 480
 DB 421 DGHCALESKVKVTCLPDGSPFSGSECHI:SGMGVTEGTGSRQLDQKLIANTLNSR 480
 QY 481 QLVDMHIDDSMICAGNLQKQDQTCGDSGGPLTCEKDTYTYVYGVISWGLECEKRPVY 540
 DB 481 QLVDMHIDDSMICAGNLQKQDQTCGDSGGPLTCEKDTYTYVYGVISWGLECEKRPVY 540
 QY 541 TQVTKFLNLWIKATIKSES GF 560
 DB 541 TQVTKFLNLWIKATIKSES GF 560

RESULT 2
 AAB42484
 ID AAB42484 standard; protein: 560 AA.
 XX
 AC AAB42484;
 XX
 DT 08-FEB-2001 (first entry)
 XX
 DE Human ORFX RFP2248 polypeptide sequence SEQ ID NO:4496.

Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
 KW vulnary; antiposiotic; antiparkinsonian; nontropic; neuroprotective;
 KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
 KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
 KW antiviral; antibacterial; antifungal; antirheumatic; antichyroid;
 KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
 KW cholesterol ester storage; systemic lupus erythematosus; infection;
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;

KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
 KW thrombosis; contraceptive.
 XX Homo sapiens.
 XX WC2000058473-A2.
 PD 05-OCT-2000.
 XX
 PF 31-MAR-2000; 2000WO-US008621.
 XX
 PR 31-MAR-1999; 99US-0127607P.
 PR 02-APR-1999; 99US-0127636P.
 PR 05-APR-1999; 99US-0127728P.
 PR 30-MAR-2000; 2000US-00540763.
 XX
 PA (CURA-) CURAGEN CORP.
 XX
 PI Shimketa RA, Leach M;
 XX
 DR WPI; 2000-602362/57.
 DR N-PSDB; AAC76693.
 XX
 PT Novel nucleic acids and peptides derived from open reading frame X,
 PT useful for treating e.g. cancers, proliferative disorders,
 PT neurodegenerative disorders and cardiovascular disease.
 XX
 PS Claim 11; Page 3681-3683; 5507pp; English.

CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
 CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
 CC sequences have activities such as: cytostatic; hepatotropic; vulnary;
 CC antiposiotic; antiparkinsonian; nontropic; neuroprotective; osteopathic;
 CC anticonvulsant; antiarthritic; immunosuppressant; antidiabetic; hypotensive;
 CC cardiant; thrombolytic; coagulant; vasotropic; antirheumatic; antichyroid;
 CC dermatological; immunosuppressive; antiinflammatory; antibacterial;
 CC antiviral; antifungal; antineumatic; antichyroid; and antianaemic. The
 CC sequences can be used for determining the presence of or predisposition
 CC to, or preventing or treating pathological conditions associated with an
 CC ORFX-associated disorder. The nucleic acids can be used to express ORFX
 CC proteins in gene therapy vectors. The proteins and nucleic acids may be
 CC used to treat cancers, proliferative disorders, neurodegenerative
 CC disorders, osteoarthritis, graft vs host disease, cardiovascular disease,
 CC diabetes mellitus, hypertension, hypothyroidism, cholesterol ester
 CC storage, systemic lupus erythematosus, severe combined immunodeficiency
 CC (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune
 CC disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and
 CC cartilage damage, nocturnal haemoglobinuria, antiinflammatory disease; to
 CC enhance coagulation; to inhibit thrombosis; and as a contraceptive

XX SQ Sequence 560 AA;
 Query Match 99.7%; Score 3143; DB 3; Length 560;
 Best Local Similarity 99.6%; Pred. No. 3.1e-185;
 Matches 558; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MFARMSDLHVLMLVKGKTACGFSLSLSLSDPDWTPDQYDYSYEDYNOEENTSSLT 60
 DB 1 MFARMSDLHVLMLVKGKTACGFSLSLSLSDPDWTPDQYDYSYEDYNOEENTSSLT 60
 QY 61 HAENPDWYTEDQADPCQPNPCBHGDCVHSGTFTCSCLAPFSGNKCQKQVNTCKDNPC 120
 DB 61 HAENPDWYTEDQADPCQPNPCBHGDCVHSGTFTCSCLAPFSGNKCQKQVNTCKDNPC 120
 QY 121 GRGCLITQSPFYRCVCKHPYTGPCSCQVVPVCRPNPCQNGATCSRHKRSKFTACPD 180
 DB 121 GRGCLITQSPFYRCVCKHPYTGPCSCQVVPVCRPNPCQNGATCSRHKRSKFTACPD 180
 QY 181 QFKGKFCIEIGSDCYVGDGYSYRGKQNRVTVNOHACLYNWSHLLLOENYMFMEDEATHGI 240
 DB 181 QFKGKFCIEIGSDCYVGDGYSYRGKQNRVTVNOHACLYNWSHLLLOENYMFMEDEATHGI 240
 QY 241 GEHNFRCNPDADEKPCWCFIKVTNDKVKWEYCDVSACSADQVAYPESPTSTKLPGPS 300

DB 241 GEHFCRNPDADKPCWCFIKVTNDKVKWEYCDVSACSAQDVAYPEESPTEPSTKLPGFDS 300
QY 301 CGKTEIAERKIKRIYGGFKSTAGKHPQASQSSLPITISMPQGHFCGGALIHPCWVLT 360
DB 301 CGKTEIAERKIKRIYGGFKSTAGKHPQASQSSLPITISMPQGHFCGGALIHPCWVLT 360
QY 361 AHCTDIKTRHLKVLVDQDLKKEEFHEQSFVQKIFKYSHYNERDEIPHNDIALLLKPV 420
DB 361 AHCTDIKTRHLKVLVDQDLKKEEFHEQSFVQKIFKYSHYNERDEIPHNDIALLLKPV 420
QY 421 DGHCALESKYVKTVCPLPDGSPFSGSECHISGWGVTETGKGRQLLDKVKLIANTLCNSR 480
DB 421 DGHCALESKYVKTVCPLPDGSPFSGSECHISGWGVTETGKGRQLLDKVKLIANTLCNSR 480
QY 481 QLYDHMIDDSMICAGNLQKPGQDTCQDGGPLTCEKDGTYVYGVISWGLECEKRPVGY 540
DB 481 QLYDHMIDDSMICAGNLQKPGQDTCQDGGPLTCEKDGTYVYGVISWGLECEKRPVGY 540
QY 541 TQVTKFLNWKIKATIKSESGF 560
DB 541 TQVTKFLNWKIKATIKSESGF 560
RESULT 3
AAO17144
ID AAO17144 standard; protein; 560 AA.
XX
AC AAO17144;
XX
DT 13-JUN-2002 (first entry)
XX
DE Human blood coagulation factor VII activating protease.
XX
KW Human; blood coagulation factor VII activating protease; FSAP;
KW single-chain plasminogen activator; bleeding disorder; haematological;
KW haemostatic.
XX
OS Homo sapiens.
XX
PN EP1182258-A1.
XX
PD 27-FEB-2002.
XX
PF 05-JUL-2001; 2001EP-00115691.
XX
PR 26-JUL-2000; 2000DE-01036641.
PR 10-OCT-2000; 2000DE-01050040.
PR 21-OCT-2000; 2000DE-01052319.
PR 12-APR-2001; 2001DE-01018706.
XX
PA (AVET) AVENTIS BEHRING GMBH.
XX
PI Roemisch J, Stoehr H, Feussner A, Lang W, Weimer T, Becker M;
PI Nerlich C, Muth-Naumann G;
XX
DR WPI; 2002-270939/32.
DR N-PSDB; AAL45696.
XX
PT New nucleic acid encoding mutant factor 7 activating protease, useful for
PT diagnosis, treatment and prevention of coagulation disorders, also
PT related protein and antibodies.
XX
PS Claim 4; Page 17-19; 27pp; German.
XX
CC The present invention relates to a mutant of the DNA sequence encoding
CC the Protease (FSAP) that activates blood coagulation factor VII (FVII)
CC and single-chain plasminogen activator, where at least one of the base
CC changes G to C at nucleotide 1177 and G to A at nucleotide 1601 is
CC present. The mutant sequences can be used in the treatment and prevention
CC of bleeding disorders associated with inherited or acquired defects of
CC blood coagulation factors V, VIII, IX, X, XI, XII, von Willebrand's
CC factor and/or with antibodies against any of these proteins. The present

CC sequence is the human FSAP protein
XX
SQ Sequence 560 AA;
Query Match 99.7%; Score 3143; DB 5; Length 560;
Best Local Similarity 99.6%; Pred. No. 3.1e-185;
Matches 558; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 MFARMSDLHLLMALVGTACGFSLSMLLESLLDPDTPDQDYDYEDYDYNQENTSSLT 60
DB 1 MFARMSDLHLLMALVGTACGFSLSMLLESLLDPDTPDQDYDYEDYDYNQENTSSLT 60
QY 61 HAENPWPYTEBOADPCQBNPCHEGGDCLVHGSTTCCLAPFSGNKKQKQVNTCKDNC 120
DB 61 HAENPWPYTEBOADPCQBNPCHEGGDCLVHGSTTCCLAPFSGNKKQKQVNTCKDNC 120
QY 121 GRQCLITQSPPYRYCVCKHPYTGPSQVQVPCPNPQNGATCSRHRKSKFTCACPD 180
DB 121 GRQCLITQSPPYRYCVCKHPYTGPSQVQVPCPNPQNGATCSRHRKSKFTCACPD 180
QY 181 QFKGKFCETGSDDCYVGDGYSYRGKXNRTVNOHACLYWNSHLLLOENTYNNFVEDAETHGI 240
DB 181 QFKGKFCETGSDDCYVGDGYSYRGKXNRTVNOHACLYWNSHLLLOENTYNNFVEDAETHGI 240
QY 241 GEHFCRNPDADKPCWCFIKVTNDKVKWEYCDVSACSAQDVAYPEESPTEPSTKLPGFDS 300
DB 241 GEHFCRNPDADKPCWCFIKVTNDKVKWEYCDVSACSAQDVAYPEESPTEPSTKLPGFDS 300
QY 301 CGKTEIAERKIKRIYGGFKSTAGKHPQASQSSLPITISMPQGHFCGGALIHPCWVLT 360
DB 301 CGKTEIAERKIKRIYGGFKSTAGKHPQASQSSLPITISMPQGHFCGGALIHPCWVLT 360
QY 361 AHCTDIKTRHLKVLVDQDLKKEEFHEQSFVQKIFKYSHYNERDEIPHNDIALLLKPV 420
DB 361 AHCTDIKTRHLKVLVDQDLKKEEFHEQSFVQKIFKYSHYNERDEIPHNDIALLLKPV 420
QY 421 DGHCALESKYVKTVCPLPDGSPFSGSECHISGWGVTETGKGRQLLDKVKLIANTLCNSR 480
DB 421 DGHCALESKYVKTVCPLPDGSPFSGSECHISGWGVTETGKGRQLLDKVKLIANTLCNSR 480
QY 481 QLYDHMIDDSMICAGNLQKPGQDTCQDGGPLTCEKDGTYVYGVISWGLECEKRPVGY 540
DB 481 QLYDHMIDDSMICAGNLQKPGQDTCQDGGPLTCEKDGTYVYGVISWGLECEKRPVGY 540
QY 541 TQVTKFLNWKIKATIKSESGF 560
DB 541 TQVTKFLNWKIKATIKSESGF 560
RESULT 4
AAR53962
ID AAR53962 standard; protein; 655 AA.
XX
AC AAR53962;
XX
DT 25-MAR-2003 (revised)
DT 06-JAN-1995 (first entry)
XX
DE Hepatocyte growth factor converting protease.
XX
KW Hepatocyte growth factor; protease; cleavage; active; inactive;
KW precursor.
XX
OS Homo sapiens.
XX
PN EP596524-A2.
XX
PD 11-MAY-1994.
XX
PF 05-NOV-1993; 93EP-00117988.
XX
PR 05-NOV-1992; 92JP-00296133.
PR 20-NOV-1992; 92JP-00312234.

```

PR 20-NOV-1992; 92JP-00312242.
XX (MITU ) MITSUBISHI KASEI CORP.
XX Shimomura T, Yamada K, Morimoto Y, Kitamura N, Miyazawa K;
XX WPT; 1994-152921/19.
XX N-PSDB; AAQ63951.
XX Hepatocyte growth factor converting protease and precursor and gene
XX encoding them - for producing active two chain HGF from inactive single
XX chain HGF.
XX Claim 12; Page 21-24; 30pp; English.
XX Hepatocyte growth factor converting protease is capable of converting
XX inactive single chain hepatocyte growth factor (HGF) into active two
XX chain HGF by cleavage at a specific site. (Updated on 25-MAR-2003 to
XX correct PN field.)
XX Sequence 655 AA;
SQ
Query Match 26.1%; Score 822.5; DB 2; Length 655;
Best Local Similarity 33.2%; Pred. No. 1.6e-42;
Matches 190; Conservative 90; Mismatches 221; Indels 71; Gaps 21;
QY 31 ESLDPDTPQDYDYSYEDYNEQENTSSLTTHAENPDW-----YTEDQA----- 74
DB 99 QALTEDEGRPCFPFRYGRMLHACTSEGSAHRK---WCATHNYDRDRAWGYCVAETPPP 155
QY 75 -----DPCQNPCEHGGDC--LVHGSTFTCCLAPFSGNKC--QKVONTCK-----D 117
DB 156 GGPAALDPCASGCPCLNGSGSCNTQDPQSYHSCSPRAFTGKDCGTEKCFDETRVEYLEGGD 215
QY 118 NPCRGQCLITQSPPYRCVKCHPYTGPSCSQV--PVCPRNPCCGATCSRHKRSKFT 175
DB 216 RWARVRQCHVEQ-----CEC---FGRTWCETGRHTRACLSPPCLNGTCHLIVAGTTV 266
QY 176 CACPDQFQKGFCEIGSD--CYVGDGYSYRGKMRITVNHQACLYWNSHLLQENYMPMED 234
DB 267 CACPPGFAGRLCNIEPDERCFLNGTGYRGVASTSASGLSCLAWNSDLLYOELHVDVSGA 326
QY 235 AETHIGIEHNCRPADDEKWCICIKVTNDKVKWEYCDVSAACA--QDVAYPEESPEPS 292
DB 327 AALLGLGPHAYCRPNPDERPWCYV-VKDSALSWEYCRLEACEBSLTVQLSPDLLATPE 385
QY 293 TKLPQFDCSGKTEIAERKIK-RYGGFKSTAGKHPQWASLQSSILPLATISMPQGHFCGAL 351
DB 386 PASPORQACGRHKKRTFLRILIGSSSLPGSHPLAAIYG-----DSFCAGSL 436
QY 352 IHPCWVLTAAHCTDIKTRH--LKVLGDQDLKKEEFHQSFVQKIFKYSHYNERDEIPH 409
DB 437 VHTCWVSAAHCFSHSPRDSVSVVLGQHFFNRITTDVTQTGIEKYIPTLYSVFNPSDH 496
QY 410 NDIALLLKVPDGHGCHALSKYKVTCLPD--GSPFSGSECHISGWG-VTEGKG-SROLL 465
DB 497 -DLVIRLKKGDRCATRSQFVQICLPEPSTFPAGHKCQIAGWHLDENVSYSLSLR 555
QY 466 DAKVKLIANTLNSQLYDHMDISMI CAGNLQKPGQDTCQDGGSPITCKDGTYYVYG 525
DB 556 EALVPLVADHKCSSPEVYGADISPNMLCAGYFDCK-SDACQDGGGPLACEKNGVAYLYG 614
QY 526 IYSWGLECEK--RPGVYTVQTKFLNWKATIK 555
DB 615 IISWGDGCGRLHKPGYTRVANYVDINDIR 646
RESULT 5
AAR89197
ID AAR89197 standard; protein; 655 AA.
XX
AC AAR89197;
XX

```

```

DT 08-AUG-1996 (first entry)
XX Human hepatocellular growth factor single chain precursor protein.
XX Mature protein; inactive; single chain; hepatocellular growth factor;
XX liver; human; precursor; dimerisation; double chain; serine protease;
XX hepatitis.
XX OS Homo sapiens.
XX Key Location/Qualifiers
XX FH 356..655
XX Peptide /note= "mature peptide"
XX JP08027026-A.
XX 30-JAN-1996.
XX 22-JUL-1994; 94JP-00171207.
XX 22-JUL-1994; 94JP-00171207.
XX (MITU ) MITSUBISHI CHEM CORP.
XX WPI; 1996-136206/14.
XX New preventative and therapeutic cpds contg. a 34 kD serine protease -
XX useful for treatment of liver diseases e.g. hepatitis.
XX Claim 4; Page 6-8; 8pp; Japanese.
XX This is the amino acid sequence of the precursor protein of the inactive
XX single chain form of a hepatocellular growth factor. The mature protein
XX (AAR89196) has a mol. wt. of around 34 kD and is derived from the 96 kD
XX precursor protein. The mature protein corresp. to residues 356-655 of the
XX precursor protein. The inactive single chain form of the growth factor is
XX activated by dimerisation of the mature protein. The active protein is a
XX serine protease which can be used for the treatment of hepatitis
XX Sequence 655 AA;
QY 31 ESLDPDTPQDYDYSYEDYNEQENTSSLTTHAENPDW-----YTEDQA----- 74
DB 99 QALTEDEGRPCFPFRYGRMLHACTSEGSAHRK---WCATHNYDRDRAWGYCVAETPPP 155
QY 75 -----DPCQNPCEHGGDC--LVHGSTFTCCLAPFSGNKCQKVONTCKDNPCGRGQL 126
DB 156 GGPAALDPCASGCPCLNGSGSCNTQDPQSYHSCSPRAFTG-----XD--CGTEKCF 203
QY 127 ITQSPPY-----RCVKHPYTGSCSQV--PVCPRNPCCGATCSRHK 169
DB 204 DETRYEYLEGGDRWARVRQGHVEQCEC---FGRTWCETGRHTRACLSPPCLNGTCHLIV 260
QY 170 RRSKFTCACPDQFKGFCBIGSD--CYVGDGYSYRGKMRITVNHQACLYWNSHLLQENY 228
DB 261 ATGTTVCACPPGFAGRLCNIEPDERCFLNGTGYRGVASTSASGLSCLAWNSDLLYOELH 320
QY 229 NMFMEADETHGIEHNCRPNPADDEKWCICIKVTNDKVKWEYCDVSAACA--QDVAYPEE 286
DB 321 VDSVGAALLGLGPHAYCRPNPDERPWCYV-VKDSALSWEYCRLEACEBSLTVQLSPDL 379
QY 287 SPTPESTKLPQFDCSGKTEIAERKIK-RYGGFKSTAGKHPQWASLQSSILPLATISMPQGH 345
DB 380 LATLPEPASGPRQACGRHKKRTFLRILIGSSSLPGSHPLAAIYG-----DS 430
QY 346 FCGGALIHPCWVLTAAHCTDIKTRH--LKVLGDQDLKKEEFHQSFVQKIFKYSHYNE 403
DB 431 FCAGSLVHTCWVSAAHCFSHSPRDSVSVVLGQHFFNRITTDVTQTGIEKYIPTLYSV 490
Query Match 26.1%; Score 821.5; DB 2; Length 655;
Best Local Similarity 32.9%; Pred. No. 1.9e-42;
Matches 190; Conservative 88; Mismatches 217; Indels 83; Gaps 21;

```

QY 124 -QCLITQPPYRCVCKHPTGSCQVVPVCEPNPCQ-----NGATCSHKRSKFTCAC 178
Db 111 AECTNWN-----SALAOKPYSGR-----RPDAIRLGLGNHNYCRNPRDRSKPWCYV 157
QY 179 PDQFK-GKF-----CBIGSDDCYVGDYSGYRGKMNRTVNOHACLYWNSHLLLOENY 228
Db 158 ---FKAGKYSSEFCSTPACSEGNSDCYFGNGSAVRGTHSLTSGASCLPWNMILIGKVY 214
QY 229 NMFMEDAETHGIGHNFCRNPDADEKFWCFIKVTNDKVKWEYCDVSAGSAQDAVPEESP 288
Db 215 TAQNPQAALGLGKHNYCRNPDGDAKFWCHV-LKNRRLTWECVDFPSCS-----262
QY 289 TEPSTKLPFGDSCGTIAERKIRIYGGFKSTAGKHPWQASLQSLPLTISMPQGHFCG 348
Db 263 -----TCGLROYSQF-RIKGGLFADIASHPWQAAIFAK---HRRSPGERFLCG 308
QY 349 GALIHPCWMLTAHCTD--IKTRHLKVLGDQDLKKEEFHQSFRVQKIFYSHYNERDE 406
Db 309 GILLSSCWILSAHCFQERPPHLLTVILGRYRVVPGEEQKFEVEKIVHKEFD--DD 366
QY 407 IPHNDIALKLKPVGDHGALESKYVTVCLP--DGSFPGSGECHISGMWVETGKG--SR 462
Db 367 TYDNDIALQLKSDSRCAQESSVVRVCLPPADLQLPDWTCELSGYGKHEALSPFYSE 426
QY 463 QLLDAKYKLIANTLNSRQLYDHMIDDSMICAGNLQKPG-----ODTCQDSDGSGPLTCEK 517
Db 427 RLKXENHVLFPSSRCTSOHLNRTVTDNMLCAGDTRSGGPOANLHDACQSDSGGLVCLN 486
QY 518 DGTYYVYGVISWGLECEK--PGVYTVTKPLNMIKATIK 555
Db 487 DGRMTLVGLIISWGLGCGQKDPGVYTKVNYLDWIRDNR 526

RESULT 14
US-09-969-271-7
; Sequence 7, Application US/09969271
; Patent No. US20020098179A1
; GENERAL INFORMATION:
; APPLICANT: Pfizer Inc. (All designated States except GB and EP (GB));
; APPLICANT: Pfizer Limited (GB and EP (GB) only);
; TITLE OF INVENTION: Pharmaceutical Combinations
; FILE REFERENCE: PCS10951APME
; CURRENT APPLICATION NUMBER: US/09/969,271
; PRIOR FILING DATE: 2001-10-01
; PRIOR APPLICATION NUMBER: GB 0025473.0
; PRIOR FILING DATE: 2000-10-17
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 562
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-969-271-7

Query Match 24.0%; Score 755.5; DB 9; Length 562;
Best Local Similarity 34.6%; Pred. No. 5.9e-51;
Matches 180; Conservative 78; Mismatches 177; Indels 85; Gaps 20;

QY 77 CQNPCEHGGDC--LVHGSTFTCTCLAPFSGNKCO-KVQNTCKDNP--CGRG-----123
Db 86 CSEPRCFNGGTCCQALYFSDVFCQCEGFGAGKCEIDTRATCYEDQGISYRGTWSTABSG 145

QY 124 -QCLITQPPYRCVCKHPTGSCQVVPVCEPNPCQ-----NGATCSHKRSKFTCAC 178
Db 146 AECTNWN-----SALAOKPYSGR-----RPDAIRLGLGNHNYCRNPRDRSKPWCYV 192

QY 179 PDQFK-GKF-----CBIGSDDCYVGDYSGYRGKMNRTVNOHACLYWNSHLLLOENY 228
Db 193 ---FKAGKYSSEFCSTPACSEGNSDCYFGNGSAVRGTHSLTSGASCLPWNMILIGKVY 249

QY 229 NMFMEDAETHGIGHNFCRNPDADEKFWCFIKVTNDKVKWEYCDVSAGSAQDAVPEESP 288
Db 250 TAQNPQAALGLGKHNYCRNPDGDAKFWCHV-LKNRRLTWECVDFPSCS-----297

QY 289 TEPSTKLPFGDSCGTIAERKIRIYGGFKSTAGKHPWQASLQSLPLTISMPQGHFCG 348
Db 298 -----TCGLROYSQF-RIKGGLFADIASHPWQAAIFAK---HRRSPGERFLCG 343
QY 349 GALIHPCWMLTAHCTD--IKTRHLKVLGDQDLKKEEFHQSFRVQKIFYSHYNERDE 406
Db 344 GILLSSCWILSAHCFQERPPHLLTVILGRYRVVPGEEQKFEVEKIVHKEFD--DD 401
QY 407 IPHNDIALKLKPVGDHGALESKYVTVCLP--DGSFPGSGECHISGMWVETGKG--SR 462
Db 402 TYDNDIALQLKSDSRCAQESSVVRVCLPPADLQLPDWTCELSGYGKHEALSPFYSE 461
QY 463 QLLDAKYKLIANTLNSRQLYDHMIDDSMICAGNLQKPG-----ODTCQDSDGSGPLTCEK 517
Db 462 RLKXENHVLFPSSRCTSOHLNRTVTDNMLCAGDTRSGGPOANLHDACQSDSGGLVCLN 521
QY 518 DGTYYVYGVISWGLECEK--PGVYTVTKPLNMIKATIK 555
Db 522 DGRMTLVGLIISWGLGCGQKDPGVYTKVNYLDWIRDNR 561

RESULT 15
US-10-411-037-26
; Sequence 26, Application US/10411037
; Publication No. US20040043446A1
; GENERAL INFORMATION:
; APPLICANT: Neose Technologies, Inc.
; APPLICANT: Defrees, Shawn
; APPLICANT: Zopf, David
; APPLICANT: Beyer, Robert
; APPLICANT: Hakes, David
; APPLICANT: Chen, Xi
; APPLICANT: Bowe, Caryn
; TITLE OF INVENTION: ALPHA GALACTOSIDASE A: REMODELING AND GLYCOCONJUGATION OF AL.
; TITLE OF INVENTION: GALACTOSIDASE A
; FILE REFERENCE: 040853-01-5082
; CURRENT APPLICATION NUMBER: US/10/411,037
; CURRENT FILING DATE: 2003-04-09
; PRIOR APPLICATION NUMBER: US 60/328,523
; PRIOR FILING DATE: 2001-10-10
; PRIOR APPLICATION NUMBER: US 60/344,692
; PRIOR FILING DATE: 2001-10-19
; PRIOR APPLICATION NUMBER: US 60/387,292
; PRIOR FILING DATE: 2002-06-07
; PRIOR APPLICATION NUMBER: US 60/391,777
; PRIOR FILING DATE: 2002-06-25
; PRIOR APPLICATION NUMBER: US 60/396,594
; PRIOR FILING DATE: 2002-07-17
; PRIOR APPLICATION NUMBER: US 60/404,249
; PRIOR FILING DATE: 2002-08-16
; PRIOR APPLICATION NUMBER: US 60/407,527
; PRIOR FILING DATE: 2002-08-28
; NUMBER OF SEQ ID NOS: 75
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 26
; LENGTH: 562
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-411-037-26

Query Match 24.0%; Score 755.5; DB 12; Length 562;
Best Local Similarity 34.6%; Pred. No. 5.9e-51;
Matches 180; Conservative 78; Mismatches 177; Indels 85; Gaps 20;

QY 77 CQNPCEHGGDC--LVHGSTFTCTCLAPFSGNKCO-KVQNTCKDNP--CGRG-----123
Db 86 CSEPRCFNGGTCCQALYFSDVFCQCEGFGAGKCEIDTRATCYEDQGISYRGTWSTABSG 145

QY 124 -QCLITQPPYRCVCKHPTGSCQVVPVCEPNPCQ-----NGATCSHKRSKFTCAC 178
Db 146 AECTNWN-----SALAOKPYSGR-----RPDAIRLGLGNHNYCRNPRDRSKPWCYV 192

NUMBER OF SEQ ID NOS: 25
 SOFTWARE: Patent in Ver. 2.1
 SEQ ID NO 19
 LENGTH: 527
 TYPE: PRT
 ORGANISM: Homo sapiens
 US-09-987-455-19

Query Match 24.0%; Score 755.5; DB 10; Length 527;
 Best Local Similarity 34.6%; Pred. No. 5.4e-51;
 Matches 180; Conservative 78; Mismatches 177; Indels 85; Gaps 20;
 77 CQNPCEHGGDC--LVHGSTFTCCCLAPPSGNKQ--KVQNTCKDNP--CGRG-----123
 51 CSFRCFNGGTCQOALYFSDVFCQPEGFAGKCEIDTRATCYEDQGISYRGTWSTAESG 110
 124 -QCLITQSPPYRYCCKHPYTGSCQVVPVCRNPQC---NGATCSRHKRSKFTCAC 178
 111 AECTNWS---SALAQKPYSGR-----RPAIRLGLGNHNYCRNPDSDSKPCVY 157
 179 PDQFK-GKF-----CEIGSDDCYVDGYSYRGKWNRTVQHACLWNSHLLQENY 228
 158 ---FKAGKYSSEFCSTPACSEGNDCYFNGSAIRGTHSLTESGASCLPWNMILIGKY 214
 229 NFMEDAEHTGIGEHNFPCRNPDADKPKCFIKVYNDKVKWEYCDVSACSADQVAYPEEP 288
 215 TAQNPQAALGLGKHNCRNPDGAKPWCHV-LKRRALTWEYCDVPSCS-----262
 289 TEPSTKLPGFDSCKXTEIAERKIKYIGFKSTAGKHPWQAS-QSSSLPTLTIIMPQGHFCG 348
 263 ---TCGLRQYSQOP-RINGGLFADIASHPWQAIAFAK--HRRSPGRFLCG 308
 349 GALIHPCWVLTAAHCTD--IKTRHLKVLGDQDLKKEEFHQSFVQKIFKYSHYNERDE 406
 309 GILISSCWILSAHCFQERPPHLLTVILGRYRVVPGEEQKFEVEKIVHKFPD--DD 366
 407 IPHNDIALKLVNLTNCRQLYDHMDDSMICAGNLQKPG-----QDTQGDGGGLTCEK 517
 427 RLKEAHVRLYPSSECTSOHLNRTVTDNMLCAGDTRSGGPOANLHDCQDGGGLVCLN 486
 518 DGTYYVYGVISWGLSECKR--PGVYTVQTKFLANWIKATIK 555
 487 DGRMTLVGIISWGLCGCKQDVPGVYTKVTNYLDWIRDNR 526

RESULT 12

US-10-432-842-1
 Sequence 1, Application US/10432842
 Publication No. US20040071707A1
 GENERAL INFORMATION:
 APPLICANT: Veronica A. CARROLL
 APPLICANT: Adrian L. HARRIS
 APPLICANT: ROY BICKNELL
 APPLICANT: Pat PRICE
 TITLE OF INVENTION: MODULATION OF CELL GROWTH
 FILE REFERENCE: 117-450 / N.79507A SER
 CURRENT APPLICATION NUMBER: US/10/432,842
 PRIOR FILING DATE: 2003-09-27
 PRIOR APPLICATION NUMBER: PCT/GB01/05244
 PRIOR FILING DATE: 2001-11-28
 PRIOR APPLICATION NUMBER: GB 0029001.5
 PRIOR FILING DATE: 2000-11-28
 NUMBER OF SEQ ID NOS: 2
 SOFTWARE: MS Word
 SEQ ID NO 1
 LENGTH: 527
 TYPE: PRT
 ORGANISM: Homo sapiens
 US-10-432-842-1

Query Match 24.0%; Score 755.5; DB 12; Length 527;
 Best Local Similarity 34.6%; Pred. No. 5.4e-51;
 Matches 180; Conservative 78; Mismatches 177; Indels 85; Gaps 20;
 77 CQNPCEHGGDC--LVHGSTFTCCCLAPPSGNKQ--KVQNTCKDNP--CGRG-----123
 51 CSFRCFNGGTCQOALYFSDVFCQPEGFAGKCEIDTRATCYEDQGISYRGTWSTAESG 110
 124 -QCLITQSPPYRYCCKHPYTGSCQVVPVCRNPQC---NGATCSRHKRSKFTCAC 178
 111 AECTNWS---SALAQKPYSGR-----RPAIRLGLGNHNYCRNPDSDSKPCVY 157
 179 PDQFK-GKF-----CEIGSDDCYVDGYSYRGKWNRTVQHACLWNSHLLQENY 228
 158 ---FKAGKYSSEFCSTPACSEGNDCYFNGSAIRGTHSLTESGASCLPWNMILIGKY 214
 229 NFMEDAEHTGIGEHNFPCRNPDADKPKCFIKVYNDKVKWEYCDVSACSADQVAYPEEP 288
 215 TAQNPQAALGLGKHNCRNPDGAKPWCHV-LKRRALTWEYCDVPSCS-----262
 289 TEPSTKLPGFDSCKXTEIAERKIKYIGFKSTAGKHPWQAS-QSSSLPTLTIIMPQGHFCG 348
 263 ---TCGLRQYSQOP-RINGGLFADIASHPWQAIAFAK--HRRSPGRFLCG 308
 349 GALIHPCWVLTAAHCTD--IKTRHLKVLGDQDLKKEEFHQSFVQKIFKYSHYNERDE 406
 309 GILISSCWILSAHCFQERPPHLLTVILGRYRVVPGEEQKFEVEKIVHKFPD--DD 366
 407 IPHNDIALKLVNLTNCRQLYDHMDDSMICAGNLQKPG-----QDTQGDGGGLTCEK 517
 427 RLKEAHVRLYPSSECTSOHLNRTVTDNMLCAGDTRSGGPOANLHDCQDGGGLVCLN 486
 518 DGTYYVYGVISWGLSECKR--PGVYTVQTKFLANWIKATIK 555
 487 DGRMTLVGIISWGLCGCKQDVPGVYTKVTNYLDWIRDNR 526

RESULT 13

US-10-360-101-203
 Sequence 203, Application US/10360101
 Publication No. US20040009550A1
 GENERAL INFORMATION:
 APPLICANT: Moll, Gert N.
 APPLICANT: Leenhouts, Cornelis J.
 TITLE OF INVENTION: Export and modification of (poly)peptide in the lantibiotic
 FILE REFERENCE: 2183-5673
 CURRENT APPLICATION NUMBER: US/10/360,101
 CURRENT FILING DATE: 2003-02-07
 PRIOR FILING DATE: 2002-05-24
 NUMBER OF SEQ ID NOS: 309
 SOFTWARE: Patent in version 3.1
 SEQ ID NO 203
 LENGTH: 527
 TYPE: PRT
 ORGANISM: Artificial Sequence
 FEATURE:
 OTHER INFORMATION: sequence of alteplase
 US-10-360-101-203

Query Match 24.0%; Score 755.5; DB 15; Length 527;
 Best Local Similarity 34.6%; Pred. No. 5.4e-51;
 Matches 180; Conservative 78; Mismatches 177; Indels 85; Gaps 20;
 77 CQNPCEHGGDC--LVHGSTFTCCCLAPPSGNKQ--KVQNTCKDNP--CGRG-----123
 51 CSFRCFNGGTCQOALYFSDVFCQPEGFAGKCEIDTRATCYEDQGISYRGTWSTAESG 110

GENERAL INFORMATION:
; APPLICANT: Chen, Hwei-Mei
; TITLE OF INVENTION: GENES EXPRESSED IN BREAST CANCER
; FILE REFERENCE: PA-0037 P
; CURRENT APPLICATION NUMBER: US/09/974,298
; CURRENT FILING DATE: 2001-10-04
; PRIOR APPLICATION NUMBER: 60/238,331
; PRIOR FILING DATE: 2000-05-10
; NUMBER OF SEQ ID NOS: 194
; SOFTWARE: PERL Program
; SEQ ID NO 145
; LENGTH: 562
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20020156263A1 1001470CD1
US-09-974-298-145

Query Match 24.0%; Score 756.5; DB 9; Length 562;
Best Local Similarity 34.6%; Pred. No. 4.9e-51;
Matches 180; Conservative 78; Mismatches 177; Indels 85; Gaps 20;
QY 77 CQNPCEHGGDC--LVHGSTFTCSCLAPFSGNKCO-KVQNTCKDNP--CGRG----- 123
Db 86 CSEPRCFNGGTCQOALYFSDVFCQPEGFAGKCCIEDTRATCYBDQGISYRGTWSTABSG 145
QY 124 -QCLITQSPPYRCVCKHPYTGSPCSQVVPVCRPNPCQ----NGATCSRHKRSKFTCAC 178
Db 146 AECTNWS---SALAQKPYSGR-----RPDAIRLGLGNHNYCRNPDSDSKPWCYV 192
QY 179 PDQFK-GKF-----CEIGSDDCYVGDGYSYRGKMTVNOHACLWNSHLLQENY 228
Db 193 ---FKAGKYSSEFCSTPACSEGNDCYFGNGSAYRGTHSLTSGASCLWNSMILIGKYV 249
QY 229 NFMWEDAETHGIGEHNFRCNPDADEKPCWCFIKVTNDKVWEYCDVSACSQAQDVAYPEESP 288
Db 250 TAQNPQAALGLGKHNYCRNPDGAKPCHV-LKNERLTWEYCDVPSCS----- 297
QY 289 TEPSTKLPGDSCGKTEIAERKIKRIYGGFKSTAGKHPQASLOSSLPLTISNPOGHFCG 348
Db 298 -----TCGLRQYSQPF-RKGGFLFADIASHPWQAIAFAK--HRRSPGERFLCG 343
QY 349 GALIHPCWLTAAHCTD--IKTRHLKVLGDQDLKKEEFHEQSFYVKIKFYSHYNERDE 406
Db 344 GILISSCWILSAACHCFQERPPPHLTVILGRTYRVVPGEBEOKFEVEKIVHKEFD--DD 401
QY 407 IPHNDIALLLKVPDGHCALESKYVTVCLP--DGSFPGSGECHISGWGVTETGKG--SR 462
Db 402 TYNDIALLLQKSDSRCAQESSVVRTVCLPPADLQLPDWTCELSGYGKHEALSPFYSE 461
QY 463 QLLDAKVKLNTLCKSRQLYDHMDSDMTCAGNLQKPG-----QDTCQDSDGGPLTCEK 517
Db 462 RLKEAHLVLPSSRCTSHLLNRTVTDNMLCAGDTRSGGPOANLHDACQDSDGGPLVCLN 521
QY 518 DGTYYVYGVISWGLECEKR--PGVYTVQTKFLNWKATIK 555
Db 522 DGRMTLVGIISWGLCGCKQDVPVYTKVTNYLDWIRDNR 561

RESULT 10
US-09-987-457-18
; Sequence 18, Application US/09987457
; Publication No. US20030013150A1
; GENERAL INFORMATION:
; APPLICANT: Manosroi, Aranya
; APPLICANT: Manosroi, Jiradej
; APPLICANT: Tayapiwatana, Chatchai
; APPLICANT: Goetz, Friedrich
; APPLICANT: Werner, Rolf-Guenther
; TITLE OF INVENTION: Methods for Large Scale Protein Production in Prokaryotes
; FILE REFERENCE: 0652.218001
; CURRENT APPLICATION NUMBER: US/09/987,457

CURRENT FILING DATE: 2001-11-14
; PRIOR APPLICATION NUMBER: 60/268,573
; PRIOR FILING DATE: 2001-02-15
; PRIOR APPLICATION NUMBER: GB 00 27 782.2
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 18
; LENGTH: 527
; TYPE: PRT
; ORGANISM: Homo sapiens (tpa)
US-09-987-457-18
Query Match 24.0%; Score 755.5; DB 10; Length 527;
Best Local Similarity 34.6%; Pred. No. 5.4e-51;
Matches 180; Conservative 78; Mismatches 177; Indels 85; Gaps 20;
QY 77 CQNPCEHGGDC--LVHGSTFTCSCLAPFSGNKCO-KVQNTCKDNP--CGRG----- 123
Db 51 CSEPRCFNGGTCQOALYFSDVFCQPEGFAGKCCIEDTRATCYEDQGISYRGTWSTABSG 110
QY 124 -QCLITQSPPYRCVCKHPYTGSPCSQVVPVCRPNPCQ----NGATCSRHKRSKFTCAC 178
Db 111 AECTNWS---SALAQKPYSGR-----RPDAIRLGLGNHNYCRNPDSDSKPWCYV 157
QY 179 PDQFK-GKF-----CEIGSDDCYVGDGYSYRGKMTVNOHACLWNSHLLQENY 228
Db 158 ---FKAGKYSSEFCSTPACSEGNDCYFGNGSAYRGTHSLTSGASCLWNSMILIGKYV 214
QY 229 NFMWEDAETHGIGEHNFRCNPDADEKPCWCFIKVTNDKVWEYCDVSACSQAQDVAYPEESP 288
Db 215 TAQNPQAALGLGKHNYCRNPDGAKPCHV-LKNERLTWEYCDVPSCS----- 262
QY 289 TEPSTKLPGDSCGKTEIAERKIKRIYGGFKSTAGKHPQASLOSSLPLTISNPOGHFCG 348
Db 263 -----TCGLRQYSQPF-RKGGFLFADIASHPWQAIAFAK--HRRSPGERFLCG 308
QY 349 GALIHPCWLTAAHCTD--IKTRHLKVLGDQDLKKEEFHEQSFYVKIKFYSHYNERDE 406
Db 309 GILISSCWILSAACHCFQERPPPHLTVILGRTYRVVPGEBEOKFEVEKIVHKEFD--DD 366
QY 407 IPHNDIALLLKVPDGHCALESKYVTVCLP--DGSFPGSGECHISGWGVTETGKG--SR 462
Db 367 TYNDIALLLQKSDSRCAQESSVVRTVCLPPADLQLPDWTCELSGYGKHEALSPFYSE 426
QY 463 QLLDAKVKLNTLCKSRQLYDHMDSDMTCAGNLQKPG-----QDTCQDSDGGPLTCEK 517
Db 427 RLKEAHLVLPSSRCTSHLLNRTVTDNMLCAGDTRSGGPOANLHDACQDSDGGPLVCLN 486
QY 518 DGTYYVYGVISWGLECEKR--PGVYTVQTKFLNWKATIK 555
Db 487 DGRMTLVGIISWGLCGCKQDVPVYTKVTNYLDWIRDNR 526

RESULT 11
US-09-987-455-19
; Sequence 19, Application US/09987455
; Publication No. US20030049729A1
; GENERAL INFORMATION:
; APPLICANT: Aranya, Manosroi
; APPLICANT: Jiradej, Manosroi
; APPLICANT: Chatchai, Tayapiwatana
; APPLICANT: Friedrich, Goetz
; APPLICANT: Rolf-Guenther, Werner
; TITLE OF INVENTION: Methods for Large Scale Production of Recombinant
; TITLE OF INVENTION: DNA-Derived tpa or K2S Molecules
; FILE REFERENCE: 0652.219001
; CURRENT APPLICATION NUMBER: US/09/987,455
; CURRENT FILING DATE: 2001-11-14
; PRIOR APPLICATION NUMBER: 60/268,574
; PRIOR FILING DATE: 2001-02-15
; PRIOR APPLICATION NUMBER: GB 0027779.8
; PRIOR FILING DATE: 2000-11-14

APPLICANT: WIEDERMANN, CHRISTIAN JOSEF
APPLICANT: ROEMISCH, JUERGEN
APPLICANT: WEIMER, THOMAS
APPLICANT: PEUSNER, ANNETTE
APPLICANT: STOEHR, HANS-ARNOLD
APPLICANT: DOERSAM, VOLKER
TITLE OF INVENTION: MAB9P I MUTANT OF FACTOR VII ACTIVATING PROTEASE
FILE REFERENCE: (PSP) AS RISK FACTOR FOR ARTERIAL THROMBOSIS
CURRENT APPLICATION NUMBER: US/10/391,215
CURRENT FILING DATE: 2003-03-19
PRIOR APPLICATION NUMBER: 09/912,559
PRIOR FILING DATE: 2001-07-26
PRIOR APPLICATION NUMBER: DE 100 36 641.4
PRIOR FILING DATE: 2000-07-26
PRIOR APPLICATION NUMBER: DE 100 50 040.4
PRIOR FILING DATE: 2000-10-10
PRIOR APPLICATION NUMBER: DE 100 52 319.6
PRIOR FILING DATE: 2000-10-21
PRIOR APPLICATION NUMBER: DE 101 18 706.8
PRIOR FILING DATE: 2001-04-12
NUMBER OF SEQ ID NOS: 8
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 5
LENGTH: 560
TYPE: PRT
ORGANISM: Homo sapiens
US-10-391-215-5

Query Match 99.7%; Score 3143; DB 15; Length 560;
Best Local Similarity 99.6%; Pred. No. 4e-239;
Matches 558; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MFARMSDLHLLMALVGTACFSLSLSLESDPDWTDPOYDYSYEDYNQENTSSLT 60
DB 1 MFARMSDLHLLMALVGTACFSLSLSLESDPDWTDPOYDYSYEDYNQENTSSLT 60
QY 61 HAENPDWYTTDQADPCQNPCEHGGDCLVHGSTFTCSCLAPFSGNKKQKVQNTCKDNPC 120
DB 61 HAENPDWYTTDQADPCQNPCEHGGDCLVHGSTFTCSCLAPFSGNKKQKVQNTCKDNPC 120
QY 121 GRGQCLITQSPPPYRCVCKHPYTGPCSCQVVPVCRPNPCNGATCSRHKRSKFTCACPD 180
DB 121 GRGQCLITQSPPPYRCVCKHPYTGPCSCQVVPVCRPNPCNGATCSRHKRSKFTCACPD 180
QY 181 QFKGKCEIGSDCYVGDGYSYRGKMRNTVNHACLYWNHSHLLQENYNNFMEDAETHGI 240
DB 181 QFKGKCEIGSDCYVGDGYSYRGKMRNTVNHACLYWNHSHLLQENYNNFMEDAETHGI 240
QY 241 GEHNFCRNPDADEKPCWCFIKVTNDKVWEYCDVSACSADQVAYPEESPSTKLPFGDS 300
DB 241 GEHNFCRNPDADEKPCWCFIKVTNDKVWEYCDVSACSADQVAYPEESPSTKLPFGDS 300
QY 301 CGKTEIAERKIKRTYGGFKSTAGHPQWASLQSSLPITISMPQGHFCGGALHPCWVLT 360
DB 301 CGKTEIAERKIKRTYGGFKSTAGHPQWASLQSSLPITISMPQGHFCGGALHPCWVLT 360
QY 361 AHCTDIKTRHLKVVLGQDQDKKEEFHEQSPRVQKIFKYSHYNERDRIPHNHIALKLKPV 420
DB 361 AHCTDIKTRHLKVVLGQDQDKKEEFHEQSPRVQKIFKYSHYNERDRIPHNHIALKLKPV 420
QY 421 DGHCALESKYVTKVCLPDGSPFSGSECHIISGWTGTGKSGRQLLDAKVKLIANTLCNSR 480
DB 421 DGHCALESKYVTKVCLPDGSPFSGSECHIISGWTGTGKSGRQLLDAKVKLIANTLCNSR 480
QY 481 QLYDHMTDDSMICAGNLQKPGQDTCQDSDGSGPLTCEKDGTYVYVGVISWGLECEKPGVY 540
DB 481 QLYDHMTDDSMICAGNLQKPGQDTCQDSDGSGPLTCEKDGTYVYVGVISWGLECEKPGVY 540
QY 541 TQVTKFLNWKATIKSSSGF 560
DB 541 TQVTKFLNWKATIKSSSGF 560

RESULT 8

US-10-172-712-28
Sequence 28, Application US/10172712
Publication No. US20030125232A1
GENERAL INFORMATION:
APPLICANT: GRIFFIN, JOHN H.
APPLICANT: GALE, ANDREW J.
APPLICANT: GETZOFF, ELIZABETH D.
APPLICANT: PELLEQUER, JEAN-LUC
TITLE OF INVENTION: STABILIZED PROTEINS WITH ENGINEERED DISULFIDE BONDS
FILE REFERENCE: 4198-4001US1
CURRENT APPLICATION NUMBER: US/10/172,712
CURRENT FILING DATE: 2002-09-30
PRIOR APPLICATION NUMBER: 60/298,578
PRIOR FILING DATE: 2001-06-14
NUMBER OF SEQ ID NOS: 32
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 28
LENGTH: 655
TYPE: PRT
ORGANISM: Homo sapiens
US-10-172-712-28

Query Match 26.1%; Score 821.5; DB 14; Length 655;
Best Local Similarity 32.9%; Pred. No. 4.5e-56;
Matches 190; Conservative 88; Mismatches 217; Indels 83; Gaps 21;

QY 31 ESIDPDWTPDOYDYSYEDYNQENTSSLTTHAENPDW-----YTEDQA----- 74
DB 99 QALTEGRPCRFPPRYGRMLHACTSGSAHRK---WCATHYVDRDRANGVCVEATPPP 155
QY 75 -----DPCPNPCEHGGDC--LVHGSTFTCSCLAPFSGNKKQKVQNTCKDNPCGRQOCL 126
DB 156 GGPAAALDPCASGCLNGSGCSNTQDPQSYHCSCPRAFTG-----XD--CGTEKCF 203
QY 127 ITQSPPPY-----RCVCKHPYTGPCSCQVVPVCRPNPCNGATCSRHK 169
DB 204 DETRYLEGGDRWARVRQGHVEQCEC---FGGTWCEGTRHTACLSCLNGTCHLIV 260
QY 170 RRSKFTCACPDQFKGKCEIGSDD-CYVGDGYSYRGKMRNTVNHACLYWNHSHLLQENY 228
DB 261 ATGTTVCACPPGFAGRLCNIEPDERCFLNGTGYRGVASTASGLSCLAWNSDLLYQELH 320
QY 229 NMFMEDAETHGIBHNFCRNPDADEKPCWCFIKVTNDKVWEYCDVSACSADQVAYPEE 286
DB 321 VDSVGAALLGLGHAYCRPNPDERPWCYV-VKDSALSWEYCRLEACESLTRVQLSPDL 379
QY 287 SPTBPTKLPGFSDCGKTEIAERKIK-RIYGGFKSTAGHPWQASLQSSLPITISMPQGH 345
DB 380 LATLPEPASPRQACGRHKKCTFLRPRIIGSSSLFGSHPLAAIYG-----DS 430
QY 346 FCGGALHPCWVLTAAHCTDIKTRH--LKVVLGQDQDKKEEFHEQSPRVQKIFKYSHYNE 403
DB 431 FCAGSLVHTCWWVSAACHFSHPSPRDSVSVVLGQHFFNRITTDVTFGIEKYPITLYSV 490
QY 404 RDETPHNDIALKLKPVGDGHCALESKYVTKVCLPD--GSPFSGSECHIISGWTGTGK 460
DB 491 FNPSDH-DELVILRLKKGDRCATRSQVQPCLEPSTFPAGHKCQIAGHGLDENVSG 549
QY 461 -SRQLLDAKVKLIANTLCNSRQLYDHMTDDSMICAGNLQKPGQDTCQDSDGSGPLTCEKDG 519
DB 550 YSSSLREALVPLVADHKCSPPEVYGADISPNMLCAGYFDCK-SDACQDSDGSGPLACENK 608
QY 520 TYVYVGVISWGLECEK--RPGVYTVQVTKFLNWKATIK 555
DB 609 VAYLYGIISWGGCGRLKPGVYTRVANYVWDNRIR 646

RESULT 9

US-09-974-298-145
Sequence 145, Application US/09974298
Patent No. US20020156263A1

APPLICANT: LANG, WIEGAND
APPLICANT: WEIMER, THOMAS
APPLICANT: BECKER, MARGRET
APPLICANT: NEELICH, CLAUDIA
APPLICANT: MUTH-NAUMANN, GUDRUN
TITLE OF INVENTION: MUTANTS OF THE FACTOR VII-ACTIVATING PROTEASE AND
DETECTION METHODS USING SPECIFIC ANTIBODIES
FILE REFERENCE: 06478.1457
CURRENT APPLICATION NUMBER: US/09/912.559
CURRENT FILING DATE: 2001-07-26
PRIOR APPLICATION NUMBER: DE 100 36 641.4
PRIOR FILING DATE: 2000-07-26
PRIOR APPLICATION NUMBER: DE 100 50 040.4
PRIOR FILING DATE: 2000-10-10
PRIOR APPLICATION NUMBER: DE 100 52 319.6
PRIOR FILING DATE: 2000-10-21
PRIOR APPLICATION NUMBER: DE 101 18 706.8
PRIOR FILING DATE: 2001-04-12
NUMBER OF SEQ ID NOS: 4
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 3
TYPE: PRT
ORGANISM: Homo sapiens
US-09-912-559-3

Query Match 99.7%; Score 3143; DB 9; Length 560;
Best Local Similarity 99.6%; Pred. No. 4e-239;
Matches 558; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MFARMSDLHVLMLALVGKTCAGFSLMSLLESLLSDPDWTPDQDYSDYEDYNOEENTSTLT 60
Db 1 MFARMSDLHVLMLALVGKTCAGFSLMSLLESLLSDPDWTPDQDYSDYEDYNOEENTSTLT 60
Qy 61 HAENPDWYTTEDQADPCQPNCEHGGDCLVHGSTFTCSCLAPFSGNKCQKQVNTCKDNPC 120
Db 61 HAENPDWYTTEDQADPCQPNCEHGGDCLVHGSTFTCSCLAPFSGNKCQKQVNTCKDNPC 120
Qy 121 GRGQCLITQSPYYRCVCKHPYTGSCQVVPVCPNCPQNGATCSRHKRSKFTACPD 180
Db 121 GRGQCLITQSPYYRCVCKHPYTGSCQVVPVCPNCPQNGATCSRHKRSKFTACPD 180
Qy 181 QFKGKFCIGSDDCYVGDGYSYRGKMNRTVNOHACLWNSHLLLOENYMFEMEDAETHGI 240
Db 181 QFKGKFCIGSDDCYVGDGYSYRGKMNRTVNOHACLWNSHLLLOENYMFEMEDAETHGI 240
Qy 241 GEHNFERNPDADKPCWCFIKVTNDVKWYCDVSACSADQVAYPEESTPSTKLPFGDS 300
Db 241 GEHNFERNPDADKPCWCFIKVTNDVKWYCDVSACSADQVAYPEESTPSTKLPFGDS 300
Qy 301 CGKTEIAERKIKRIYGGFKSTAGKHPWQASLQSSPLTIISMPQGHFCGALIHPCWVLT 360
Db 301 CGKTEIAERKIKRIYGGFKSTAGKHPWQASLQSSPLTIISMPQGHFCGALIHPCWVLT 360
Qy 361 AHCTDIKTRHLKVVLGDQDLKKEEFHQSFRVKIFKYSHYNERDEIPHNDIALLKLPV 420
Db 361 AHCTDIKTRHLKVVLGDQDLKKEEFHQSFRVKIFKYSHYNERDEIPHNDIALLKLPV 420
Qy 421 DGHCALESKYVKTVCPLPDGSPFSGSECHISGWSGTETGKSRQLLDKVKLIANTLCNSR 480
Db 421 DGHCALESKYVKTVCPLPDGSPFSGSECHISGWSGTETGKSRQLLDKVKLIANTLCNSR 480
Qy 481 QLYDHMIDDSMICAGNLQKPGQDTCCGDSGGPLTCEKDGTYVYVGVISWGLECEKRPVY 540
Db 481 QLYDHMIDDSMICAGNLQKPGQDTCCGDSGGPLTCEKDGTYVYVGVISWGLECEKRPVY 540
Qy 541 TQVTKFLNWKATIKSESGF 560
Db 541 TQVTKFLNWKATIKSESGF 560

RESULT 6
US-10-172-712-32

Sequence 32, Application US/10172712
Publication No. US20030125232A1
GENERAL INFORMATION:
APPLICANT: GRIFFIN, JOHN H.
APPLICANT: GALE, ANDREW J.
APPLICANT: GEIZOFF, ELIZABETH D.
APPLICANT: PELUEQUER, JEAN-LUC
TITLE OF INVENTION: STABILIZED PROTEINS WITH ENGINEERED DISULFIDE BONDS
FILE REFERENCE: 4198-4001US1
CURRENT APPLICATION NUMBER: US/10/172.712
CURRENT FILING DATE: 2002-09-30
PRIOR APPLICATION NUMBER: 60/298,578
PRIOR FILING DATE: 2001-06-14
NUMBER OF SEQ ID NOS: 32
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 32
LENGTH: 560
TYPE: PRT
ORGANISM: Homo sapiens
US-10-172-712-32

Query Match 99.7%; Score 3143; DB 14; Length 560;
Best Local Similarity 99.6%; Pred. No. 4e-239;
Matches 558; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MFARMSDLHVLMLALVGKTCAGFSLMSLLESLLSDPDWTPDQDYSDYEDYNOEENTSTLT 60
Db 1 MFARMSDLHVLMLALVGKTCAGFSLMSLLESLLSDPDWTPDQDYSDYEDYNOEENTSTLT 60
Qy 61 HAENPDWYTTEDQADPCQPNCEHGGDCLVHGSTFTCSCLAPFSGNKCQKQVNTCKDNPC 120
Db 61 HAENPDWYTTEDQADPCQPNCEHGGDCLVHGSTFTCSCLAPFSGNKCQKQVNTCKDNPC 120
Qy 121 GRGQCLITQSPYYRCVCKHPYTGSCQVVPVCPNCPQNGATCSRHKRSKFTACPD 180
Db 121 GRGQCLITQSPYYRCVCKHPYTGSCQVVPVCPNCPQNGATCSRHKRSKFTACPD 180
Qy 181 QFKGKFCIGSDDCYVGDGYSYRGKMNRTVNOHACLWNSHLLLOENYMFEMEDAETHGI 240
Db 181 QFKGKFCIGSDDCYVGDGYSYRGKMNRTVNOHACLWNSHLLLOENYMFEMEDAETHGI 240
Qy 241 GEHNFERNPDADKPCWCFIKVTNDVKWYCDVSACSADQVAYPEESTPSTKLPFGDS 300
Db 241 GEHNFERNPDADKPCWCFIKVTNDVKWYCDVSACSADQVAYPEESTPSTKLPFGDS 300
Qy 301 CGKTEIAERKIKRIYGGFKSTAGKHPWQASLQSSPLTIISMPQGHFCGALIHPCWVLT 360
Db 301 CGKTEIAERKIKRIYGGFKSTAGKHPWQASLQSSPLTIISMPQGHFCGALIHPCWVLT 360
Qy 361 AHCTDIKTRHLKVVLGDQDLKKEEFHQSFRVKIFKYSHYNERDEIPHNDIALLKLPV 420
Db 361 AHCTDIKTRHLKVVLGDQDLKKEEFHQSFRVKIFKYSHYNERDEIPHNDIALLKLPV 420
Qy 421 DGHCALESKYVKTVCPLPDGSPFSGSECHISGWSGTETGKSRQLLDKVKLIANTLCNSR 480
Db 421 DGHCALESKYVKTVCPLPDGSPFSGSECHISGWSGTETGKSRQLLDKVKLIANTLCNSR 480
Qy 481 QLYDHMIDDSMICAGNLQKPGQDTCCGDSGGPLTCEKDGTYVYVGVISWGLECEKRPVY 540
Db 481 QLYDHMIDDSMICAGNLQKPGQDTCCGDSGGPLTCEKDGTYVYVGVISWGLECEKRPVY 540
Qy 541 TQVTKFLNWKATIKSESGF 560
Db 541 TQVTKFLNWKATIKSESGF 560

RESULT 7
US-10-391-215-5
Sequence 5, Application US/10391215
Publication No. US20040009543A1
GENERAL INFORMATION:
APPLICANT: KIECHL, STEFAN
APPLICANT: WILLEIT, JOHANN

; PRIOR APPLICATION NUMBER: DE 100 52 319.6
; PRIOR FILING DATE: 2000-10-21
; PRIOR APPLICATION NUMBER: DE 101 18 706.8
; PRIOR FILING DATE: 2001-04-12
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 560
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-391-215-6

Query Match 99.9%; Score 3150; DB 15; Length 560;
Best Local Similarity 99.8%; Pred. No. 1.1e-239;
Matches 559; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MFARMSDLHVLMLALVGTACGFSLSLLESLLSDPDWTPDQDYSDYEDYNOEBNTSSTLT 60
DB 1 MFARMSDLHVLMLALVGTACGFSLSLLESLLSDPDWTPDQDYSDYEDYNOEBNTSSTLT 60

QY 61 HAENPDWYTTEDQADPCQPNPCHEHGGDCLVHGSTFTCSCLAPFSGNKCQKQNTCKDNPC 120
DB 61 HAENPDWYTTEDQADPCQPNPCHEHGGDCLVHGSTFTCSCLAPFSGNKCQKQNTCKDNPC 120

QY 121 GRGQCLITQSPPYRCVCKHPYTGSPSCQVVPVCRPNPCQNGATCSRHKRSKFTCACPD 180
DB 121 GRGQCLITQSPPYRCVCKHPYTGSPSCQVVPVCRPNPCQNGATCSRHKRSKFTCACPD 180

QY 181 QFKGFCEIGSDDCVVGDSYRGKQNRVTNQHACLWNSHLLLOENYMFMEDEATHGI 240
DB 181 QFKGFCEIGSDDCVVGDSYRGKQNRVTNQHACLWNSHLLLOENYMFMEDEATHGI 240

QY 241 GEHNFRCNPDADEKPCWCFIKVTNDKVWEYCDVSACSAQDVAYPEESTPSTKLPGFDS 300
DB 241 GEHNFRCNPDADEKPCWCFIKVTNDKVWEYCDVSACSAQDVAYPEESTPSTKLPGFDS 300

QY 301 CGKTEIAERKIKRIYGGFKSTAGKHPWQASLOSPLTISMPQGHFCGGLIHPCWVLT 360
DB 301 CGKTEIAERKIKRIYGGFKSTAGKHPWQASLOSPLTISMPQGHFCGGLIHPCWVLT 360

QY 361 AHCTDIKTRHLKVVLGDQDLKKEEFHEQSFVRVOKIFKYSHYNERDEIPHNDIALKLKPV 420
DB 361 AHCTDIKTRHLKVVLGDQDLKKEEFHEQSFVRVOKIFKYSHYNERDEIPHNDIALKLKPV 420

QY 421 DGHCALESKYVKTVCPLPDGSPGSECHISGNGVTETGKSRQLLDKVKLIANTLCNSR 480
DB 421 DGHCALESKYVKTVCPLPDGSPGSECHISGNGVTETGKSRQLLDKVKLIANTLCNSR 480

QY 481 QLYDHMIDDSMICAGNLQKPGQDTCCGDSGGPLTCEKDGTYVYVGVISWGLECEKRPVY 540
DB 481 QLYDHMIDDSMICAGNLQKPGQDTCCGDSGGPLTCEKDGTYVYVGVISWGLECEKRPVY 540

QY 541 TQVTKFLNWKATIKSES GF 560
DB 541 TQVTKFLNWKATIKSES GF 560

RESULT 4
US-10-391-215-7
; Sequence 7, Application US/10391215
; Publication No. US20040009543A1
; GENERAL INFORMATION:
; APPLICANT: KIECHL, STEFAN
; APPLICANT: WILLEIT, JOHANN
; APPLICANT: WIEDERMANN, CHRISTIAN JOSEF
; APPLICANT: ROEMISCH, JUERGEN
; APPLICANT: WEIMER, THOMAS
; APPLICANT: FEUSSNER, ANNETTE
; APPLICANT: STOEHR, HANS-ARNOLD
; APPLICANT: DOERSAM, VOLKER
; TITLE OF INVENTION: MARBURG I MUTANT OF FACTOR VII ACTIVATING PROTEASE
; TITLE OF INVENTION: (FSAP) AS RISK FACTOR FOR ARTERIAL THROMBOSIS
; FILE REFERENCE: 06478.1457-01

; CURRENT APPLICATION NUMBER: US/10/391,215
; CURRENT FILING DATE: 2003-03-19
; PRIOR APPLICATION NUMBER: 09/912,559
; PRIOR FILING DATE: 2001-07-26
; PRIOR APPLICATION NUMBER: DE 100 36 641.4
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: DE 100 50 040.4
; PRIOR FILING DATE: 2000-10-10
; PRIOR APPLICATION NUMBER: DE 100 52 319.6
; PRIOR FILING DATE: 2000-10-21
; PRIOR APPLICATION NUMBER: DE 101 18 706.8
; PRIOR FILING DATE: 2001-04-12
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 7
; LENGTH: 560
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-391-215-7

Query Match 99.8%; Score 3146; DB 15; Length 560;
Best Local Similarity 99.8%; Pred. No. 2.3e-239;
Matches 559; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MFARMSDLHVLMLALVGTACGFSLSLLESLLSDPDWTPDQDYSDYEDYNOEBNTSSTLT 60
DB 1 MFARMSDLHVLMLALVGTACGFSLSLLESLLSDPDWTPDQDYSDYEDYNOEBNTSSTLT 60

QY 61 HAENPDWYTTEDQADPCQPNPCHEHGGDCLVHGSTFTCSCLAPFSGNKCQKQNTCKDNPC 120
DB 61 HAENPDWYTTEDQADPCQPNPCHEHGGDCLVHGSTFTCSCLAPFSGNKCQKQNTCKDNPC 120

QY 121 GRGQCLITQSPPYRCVCKHPYTGSPSCQVVPVCRPNPCQNGATCSRHKRSKFTCACPD 180
DB 121 GRGQCLITQSPPYRCVCKHPYTGSPSCQVVPVCRPNPCQNGATCSRHKRSKFTCACPD 180

QY 181 QFKGFCEIGSDDCVVGDSYRGKQNRVTNQHACLWNSHLLLOENYMFMEDEATHGI 240
DB 181 QFKGFCEIGSDDCVVGDSYRGKQNRVTNQHACLWNSHLLLOENYMFMEDEATHGI 240

QY 241 GEHNFRCNPDADEKPCWCFIKVTNDKVWEYCDVSACSAQDVAYPEESTPSTKLPGFDS 300
DB 241 GEHNFRCNPDADEKPCWCFIKVTNDKVWEYCDVSACSAQDVAYPEESTPSTKLPGFDS 300

QY 301 CGKTEIAERKIKRIYGGFKSTAGKHPWQASLOSPLTISMPQGHFCGGLIHPCWVLT 360
DB 301 CGKTEIAERKIKRIYGGFKSTAGKHPWQASLOSPLTISMPQGHFCGGLIHPCWVLT 360

QY 361 AHCTDIKTRHLKVVLGDQDLKKEEFHEQSFVRVOKIFKYSHYNERDEIPHNDIALKLKPV 420
DB 361 AHCTDIKTRHLKVVLGDQDLKKEEFHEQSFVRVOKIFKYSHYNERDEIPHNDIALKLKPV 420

QY 421 DGHCALESKYVKTVCPLPDGSPGSECHISGNGVTETGKSRQLLDKVKLIANTLCNSR 480
DB 421 DGHCALESKYVKTVCPLPDGSPGSECHISGNGVTETGKSRQLLDKVKLIANTLCNSR 480

QY 481 QLYDHMIDDSMICAGNLQKPGQDTCCGDSGGPLTCEKDGTYVYVGVISWGLECEKRPVY 540
DB 481 QLYDHMIDDSMICAGNLQKPGQDTCCGDSGGPLTCEKDGTYVYVGVISWGLECEKRPVY 540

QY 541 TQVTKFLNWKATIKSES GF 560
DB 541 TQVTKFLNWKATIKSES GF 560

RESULT 5
US-09-912-559-3
; Sequence 3, Application US/09912559
; Patent No. US20020142316A1
; GENERAL INFORMATION:
; APPLICANT: ROEMISCH, JUERGEN
; APPLICANT: STOEHR, HANS-ARNOLD
; APPLICANT: FEUSSNER, ANNETTE

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 24, 2004, 09:49:06 ; Search time 42 Seconds
(without alignments)
3718.956 Million cell updates/sec

Title: US-09-912-559-4

Perfect score: 3153

Sequence: 1 MFARMSDLHVLMLVAGKT.....TQVTKFLNWKATIKSBSGF 560

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1149313 seqs, 278921704 residues

Total number of hits satisfying chosen parameters: 1149313

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA.*

- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/1/pubpaa/PTCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/1/pubpaa/PTCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3153	100.0	560	9	US-09-912-559-4
2	3153	100.0	560	15	US-10-391-215-8
3	3150	99.8	560	15	US-10-391-215-6
4	3146	99.8	560	15	US-10-391-215-7
5	3143	99.7	560	9	US-09-912-559-3
6	3143	99.7	560	14	US-10-172-712-32
7	3143	99.7	560	15	US-10-391-215-5
8	821.5	26.1	655	14	US-10-172-712-28
9	756.5	24.0	562	9	US-09-974-298-145
10	755.5	24.0	527	10	US-09-987-457-18
11	755.5	24.0	527	10	US-09-987-455-19
12	755.5	24.0	527	12	US-10-432-842-1
13	755.5	24.0	527	15	US-10-360-101-203
14	755.5	24.0	562	9	US-09-969-271-7
15	755.5	24.0	562	12	US-10-411-037-26

16	755.5	24.0	562	12	US-10-411-026-26
17	755.5	24.0	562	14	US-10-193-656-8
18	755.5	24.0	562	14	US-10-443-701-4
19	755.5	24.0	562	16	US-10-410-962-26
20	755.5	24.0	562	16	US-10-411-049-26
21	742.5	23.5	650	15	US-10-401-077-1
22	733	23.2	615	10	US-09-858-909-2
23	733	23.2	615	14	US-10-172-712-30
24	733	23.2	615	16	US-10-449-132-2
25	708	22.5	372	9	US-09-084-491A-3
26	708	22.5	372	13	US-10-102-704-3
27	698.5	22.2	354	10	US-09-987-457-10
28	698.5	22.2	354	10	US-09-987-455-11
29	698.5	22.2	377	10	US-09-987-455-8
30	677	21.5	433	12	US-10-087-192-591
31	660.5	20.9	411	15	US-10-407-821-2
32	659.5	20.9	411	9	US-09-880-503-3
33	659.5	20.9	431	12	US-10-411-037-34
34	659.5	20.9	431	12	US-10-411-026-34
35	659.5	20.9	431	13	US-10-076-421-2
36	659.5	20.9	431	14	US-10-171-311-184
37	659.5	20.9	431	14	US-10-301-822-161
38	659.5	20.9	431	14	US-10-131-985-21
39	659.5	20.9	431	15	US-10-295-027-414
40	659.5	20.9	431	15	US-10-295-027-1275
41	659.5	20.9	431	16	US-10-410-962-34
42	659.5	20.9	431	16	US-10-411-049-34
43	659.5	20.9	437	12	US-10-087-192-594
44	659	20.9	812	12	US-09-825-751A-71
45	658.5	20.9	431	14	US-10-247-671-149

ALIGNMENTS

RESULT 1

US-09-912-559-4

Sequence 4, Application US/09912559

Patent No. US20020142316A1

GENERAL INFORMATION:

APPLICANT: ROEMISCH, JUERGEN

APPLICANT: STOEHR, HANS-ARNOLD

APPLICANT: FEUSSNER, ANNETTE

APPLICANT: LANG, WIEGAND

APPLICANT: WEIMER, THOMAS

APPLICANT: BECKER, MARGRET

APPLICANT: NESLICH, CLAUDIA

APPLICANT: MUTH-NAUMANN, GUDRUN

TITLE OF INVENTION: MUTANTS OF THE FACTOR VII-ACTIVATING PROTEASE AND

DETECTION METHODS USING SPECIFIC ANTIBODIES

FILE REFERENCE: 06478.1457

CURRENT APPLICATION NUMBER: US/09/912.559

PRIOR FILING DATE: 2001-07-26

PRIOR APPLICATION NUMBER: DE 100 36 641.4

PRIOR FILING DATE: 2000-07-26

PRIOR APPLICATION NUMBER: DE 100 50 040.4

PRIOR FILING DATE: 2000-10-10

PRIOR APPLICATION NUMBER: DE 100 52 319.6

PRIOR FILING DATE: 2000-10-21

PRIOR APPLICATION NUMBER: DE 101 18 706.8

PRIOR FILING DATE: 2001-04-12

NUMBER OF SEQ ID NOS: 4

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 4

LENGTH: 560

TYPE: PRT

ORGANISM: Homo sapiens

US-09-912-559-4

Query Match 100.0%; Score 3153; DB 9; Length 560;

Best Local Similarity 100.0%; Pred. No. 6.5e-240;

Matches 560; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Search completed: May 24, 2004, 09:58:58
Job time : 43 secs